CAN YOUNGER AND OLDER ADULTS JUDGE THE QUALITY OF THEIR TEXT SUMMARIES?

A Dissertation
Presented to
The Academic Faculty

By

Erika K. Fulton

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Psychology

Georgia Institute of Technology

August, 2015

CAN YOUNGER AND OLDER ADULTS JUDGE THE QUALITY OF THEIR TEXT SUMMARIES?

Approved by:

Dr. Christopher Hertzog, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Frank Durso
School of Psychology
Georgia Institute of Technology

Dr. Paul Verhaeghen
School of Psychology
Georgia Institute of Technology

Dr. John Dunlosky
Department of Psychology
Kent State University

Dr. Ute Fischer
School of Literature, Media, and Communication
Georgia Institute of Technology

Date Approved: April 17, 2015

ACKNOWLEDGEMENTS

Thank you to the members of my committee, particularly my advisor, Dr. Christopher Hertzog, for their time and input. I would also like to acknowledge the contribution of our study participants and the following undergraduate research assistants: Margeaux Comerford, Stephanie Kasper, Sujal Desai, Thomas Loalbo, Zurain Hassan, Daniel Litowitz, Raghav Srinath, Jennifer Rattray, Ceara Baage, Angelique Soulakos, Cole Alford, Kennedy Singleton, Sadiya Dhanani, Gem Valencia, Cambre Mabry, and especially Jason Chan, Kevin Pan, and Raine Hayes. This material is based upon work supported by the Ruth L. Kirschstein Training Grant, "Research Training in Cognitive Aging," (NIA T32 AG00175), as well as two grants awarded to Erika Fulton, an APA Dissertation Award, and the APF/CogDop William C. Howell Scholarship. I would also like to thank my family and friends who supported me throughout graduate school.

TABLE OF CONTENTS

LIST OF TABLES

SUMMARY

The ability to judge expository text summary quality is relevant to education, social information exchange, and important work and life decisions. Students need to judge whether they are prepared to summarize what they have read for an exam and whether their essay responses are adequate. All adults need to be able to accurately judge whether they are prepared to summarize written materials to a boss or colleague who may make important decisions based on that summary. Even effectively assessing the quality of our summaries to acquaintances and strangers is essential if we wish not to spread inaccurate information upon which others' decisions could be wrongly based. Despite the importance of the ability to judge text comprehension and predict summary quality, relatively little research has been done in this area, particularly among adults. Summaries have rarely been used to study young adult metacomprehension (Schommer & Surber, 1986) and never to study aging and metacomprehension. Although various cognitive abilities have been shown to decline with age (Verhaeghen & Salthouse, 1997), findings are mixed on whether this pattern of decline characterizes text comprehension (Hultsch, Hertzog, Dixon, & Small, 1998; Verhaeghen, Marcoen, & Goossens, 1993) and metacomprehension (Baker, Dunlosky, & Hertzog, 2010; Olin & Zelinski, 1997). Older adult text comprehension has rarely been measured with text summaries; findings vary with both age-related declines (Adams, 1991; Byrd, 1985) and improvements (Jackson & Kemper, 1993) in text recall reported. Given that different reading goals (Stine-Morrow, Shake, Miles, & Noh, 2006; Van den Broek, Lorch, Linderholm, & Gustafson, 2001) and working memory capacity (Kintsch & van Dijk, 1978) may affect how text is processed and what is comprehended from it, it is worth examining whether people are aware of

these relationships and how the awareness might affect their summaries and summary judgments. Thus, the dissertation assessed the ability to judge one's summarizing ability, and whether age, reading goals, or working memory, affect this ability. Research questions were addressed using traditional metacognitive analyses and multi-level modeling (MLM).

Participants showed some metacomprehension ability, as measured with summaries, but with much room for improvement. According to traditional statistics used in metacognition studies, older adults were more often overconfident than younger adults, with comparatively greater age equivalency in the ability to discriminate among passages more or less well understood. Multilevel modeling (MLM) suggested a pattern marked by individual differences, better between-person than within-person accuracy, and more age equivalency. MLM also suggested a more important influence of working memory than did traditional statistics. Furthermore, although reading goal moderated some age effects, as measured by traditional analyses, its effects were largely independent of age, as measured by MLM. Although the need for improved metacomprehension, particularly for older adults, is apparent, there is reason to believe this can be done without intensive training, as task experience alone induced more accurate online judgments and more consistency between general subjective and objective metacomprehension abilities.

CHAPTER 1

INTRODUCTION

Text Comprehension vs. Text Recall

According to van Dijk and Kintsch (1983), truly understanding text requires construction of a situation model. This model represents understanding of the situation described by the text that goes beyond the words and their relationship to each other, including a continuous updating of meaning that requires integration of current text meaning with the reader's knowledge and previous text content. During reading, the meanings of decoded words and their associates are automatically activated to construct the textbase, independently of the discourse context. The reader's knowledge is subsequently used to edit the textbase to be consistent with the discourse context. Comprehension of a text develops over repeated cycles of construction and integration in a relatively automatic manner, unless coherence cannot be established (Kintsch, 1988).

Text comprehension is often confused with text recall. That which is remembered is not necessarily all that is understood; and understanding is not a requirement for recall (Durso, Rawson, & Girotto, 2007; Kintsch, 1994). One can recite a poem, demonstrating good memory, without having understood it. Indeed, many researchers purporting to study comprehension have assessed it with tests that largely measure recognition memory or recall of text content. Although that which is comprehended may be more likely to be remembered, text comprehension and text memory are not the same construct (Kintsch, 1994).

Tested text comprehension may not capture all that is actually comprehended. Although some multiple choice tests in text studies include questions that require

participants to draw inferences from implicit information, and are thus a better test of true comprehension, they do not allow participants to demonstrate comprehension that was not probed. That is to say, multiple choice tests do not usually ask about every piece of information implied by the text and may underestimate what the reader has learned. They also do not tell us anything about what the reader deems most important, which is an important aspect of comprehension.

Measuring Text Comprehension

In lieu of test questions, participants can be asked to freely recall or summarize text to demonstrate what they may have understood. These methods do not limit the participants to what the experimenter chooses to query, but they confound comprehension and production—readers may understand more than they choose to or are able to express when instructed. In reality, free recall or summarizing and answering test questions about a text all contain a mixture of what is represented by the reader at the textbase level *and* the situation model (Kintsch, 1994). Given the pros and cons to the more and less constraining assessments, and the fact that no measure is a pure assessment of comprehension, it has been recommended that both be used whenever possible (Taylor, 1984).

Soliciting summaries as a measure of comprehension has its advantages. Summarizing shows whether a participant understands the parts of a text that are more important than others, albeit the quality of the summary is partly dependent on the summarizer's ability to organize and express thoughts (Garner, 1985; Kintsch & Kozminsky, 1977; Taylor, 1984). Although readers may output less when summarizing than during free recall (van Dijk, 1979), there is evidence that the number of details

provided at test under either instruction is the same, with more main ideas being represented in summaries (Riley & Lee, 1996). Whereas free recall instructions result in more literal recall, summarization instructions bring about more inferences (Vieiro & García-Madruga, 1997) especially under oral report conditions (Kellogg, 2007). Thus, summarization instructions appear to bring about more of the quality (inferences and main ideas) that would match expectations for situation model construction than free recall instructions. However, recent pilot testing in our lab indicates that for some individuals "summarize" is almost synonymous with "free recall," especially in testing situations, whereas for others it holds the intended meaning to provide only the main points and important details in an organized fashion (Brown & Day, 1983; Kintsch & van Dijk, 1978). But, with clear instructions provided to specify the intended meaning of "summarize," summaries should provide a measure of comprehension as defined by situation model construction that is similar to inference based multiple choice questions and gist level proposition recall, but which also reveals that which the reader perceives as most important.

The Role of Reading Goals

Reading goals may affect how texts are processed and understood. Younger adults report reading school texts more carefully than non-school texts (Lorch, Lorch, & Klusewitz, 1993), and report more rereading when given a study goal than an entertain goal (Narvaez, van den Broek, & Ruiz, 1999), but it is not known whether they actually do. According to think-aloud reports, there does not seem to be a difference in how deeply younger adults process texts for a study or entertain goal (Kendeou, Bohn-Gettler, & Fulton, 2011). However, when asked to recall what they read, there is more

paraphrasing in the study condition and more evaluations and associations in the entertain condition (Linderholm & van den Broek, 2002; Van den Broek et al., 2001).

Study goals sometimes result in better memory of text (Van den Broek et al., 2001) but not always (Linderholm, Cong, & Zhao, 2008). When asked to retell a story for accuracy vs. for entertainment, younger adults are more accurate and detailed in the former, although multiple choice comprehension tests reveal no difference between the conditions (Dudukovic, Marsh, & Tversky, 2004). Thus the reading goal may affect how text is processed and appears to affect what is output at test, with different test formats leading to different conclusions about the nature of readers' comprehension and how it is influenced by the reading goal.

Text Comprehension in Older Adulthood

The Role of Vocabulary and Education

The effect of age on text comprehension appears to be largely mediated by age-related variables, but results are fairly mixed on the nature of these effects. In other words, there may be variables that are correlated with both age and text comprehension that can account for a significant portion of the relationship *between* age and text comprehension. Three regularly cited mediator variables are verbal ability (Meyer & Rice, 1989), education (Hultsch, Hertzog, & Dixon, 1990), and working memory (Bryan & Luszcz, 1999; De Beni, Borella, & Carretti, 2007; DeDe, Caplan, Kemtes, & Waters, 2004; Dijkstra, 2001; Stine-Morrow et al., 2006; Van der Linden et al., 1999), but the findings do not afford definitive conclusions. Higher older adult vocabulary scores have been shown to be associated with comparable text recall (Harker, Hartley, & Walsh, 1982) but also poorer comprehension (Hartley, 1986, 1993b), even when age groups are

equated on working memory (Hartley, 1993b). When age groups are matched on years of education, however, vocabulary can still have an effect on age-related differences in text comprehension (Meyer & Rice, 1989). But vocabulary has also been shown to not explain the differences above and beyond more basic cognitive abilities (Hultsch et al., 1990) and has been questioned as a reliable predictor of text memory and as a good index of general verbal ability (Hartley, 1993a). Furthermore, although uneducated younger adults have been found to perform more similarly to older adults than to educated younger adults (Ratner, Schell, Crimmins, Mittelman, & Baldinelli, 1987), a meta-analysis suggests greater older adult education does not ensure they perform more comparably to younger adults (Johnson, 2003).

The Role of Working Memory Capacity

The role of working memory capacity in explaining age-related differences in text comprehension is unclear, in part because working memory capacity may affect parts of the reading process but not always comprehension itself. Although better working memory has been shown to predict text recall of narratives (Hultsch et al., 1990; Radvansky & Copeland, 2004), work in younger adults has shown that working memory measures are related to textbase level processing but not to situation model comprehension (Radvansky & Copeland, 2004). In line with this finding, older adults with lower working memory capacity spend more time processing words (Stine-Morrow, Soederberg Miller, Gagne, & Hertzog, 2008) than older adults with higher working memory capacity, and may have greater difficulty processing texts than younger adults, but they do not always comprehend texts less well. Indeed, although propositional recall (Radvansky, 1999) and recognition (Radvansky, Zwaan, Curiel, & Copeland, 2001)

decline with age, there is evidence that the ability to form a situational model does not (Madden & Dijkstra, 2010; Radvansky & Dijkstra, 2007). However, representing a situation model was measured by whether one confused two sentences that potentially describe the same situation more than two sentences that could not describe the same situation. Dijkstra and colleagues found that the former was confused by both age groups more than the latter.

Stine-Morrow and colleagues have shown that demonstrating an effect of working memory on older adult text recall/comprehension depends on a variety of factors. Due to reduced working memory capacity, older adults tend to allocate their resources differently, either avoiding the taxing process of conceptual integration at the ends of sentences (wrap-up; Stine, 1990) or by spending more time at clause boundaries to achieve comparable recall (Stine-Morrow et al., 2008). However, sometimes devoting the same amount of time to sentence processing as do younger adults does not equally benefit older adults' recall (Stine-Morrow et al., 2006). Despite reduced working memory, older adults can also reach comparable levels of text recall by reading more slowly (Stine-Morrow, Milinder, Pullara, & Herman, 2001; Stine & Hindman, 1994), having greater domain knowledge (Soederberg Miller, 2009) or print exposure (Payne, Gao, Noh, Anderson, & Stine-Morrow, 2012), or by developing a schema early on that they can use to facilitate subsequent textbase processing (Stine-Morrow, Loveless, & Soederberg, 1996). Despite the many ways in which older adults can compensate for deficient working memory, higher working memory capacity still affords better recall (Stine-Morrow et al., 1996) and those older adults with higher working memory capacity can spend more time at wrap-up to their benefit (Payne et al., 2012).

Other work supports that by Stine-Morrow and her colleagues regarding the role of working capacity in older adult text comprehension. Other labs have shown that texts vary in ways that may not exceed older adult working memory capacity (Radvansky, Gerard, Zacks, & Hasher, 1990), that older adults can apply schemata to constrain the effort needed (Radvansky & Dijkstra, 2007), or that older adults reread or read more slowly to compensate (Brébion, 2003; Liu, Kemper, & Bovaird, 2009). As Johnson's (2003) meta-analysis suggests, working memory may be found to be a significant mediator to the extent that the interval between reading and recall is short (1-20 minutes), as this is the interval around which age-related differences in text recall are more often found. In summary, although parts of the reading process are relatively automatic (Kintsch & van Dijk, 1978), the construction of situation models taxes working memory but to varying degrees dependent on reader and text characteristics (Radvansky & Copeland, 2004).

As suggested by some previous studies, working memory deficits may be relatively inconsequential to text comprehension if the reader can resolve comprehension difficulties by rereading. However, the effect access to previous text may have on text comprehension has not been tested systematically. In fact, access to previous text is often treated as a nuisance variable and is experimentally controlled (Jackson & Kemper, 1993). Hartley, Stojack, Mushaney, Annon, and Lee (1994) showed that older adults, without access to prior text, had poorer prose recall even though their recall of single sentences was matched to that of younger adults. The difference, the authors speculated, may have been that integration across multiple sentences when asked to read and recall prose was impaired. Access to previous text allows the reinstatement of information that

was forgotten or never encoded and not having it may have impaired integration. Comparing recall of written text to recall of heard text supports this conclusion. Taub and Kline (1976) found that text recall is worse for auditory than written text (in either age group) *only* when participants in the reading condition were allowed to look back in the written text. Indeed, Johnson's (2003) meta-analysis revealed that smaller age differences in text recall are found when text is written as opposed to heard, likely because only when text is written can one return to previous information.

Although presenting sentences one at a time with no opportunity to look back affords a measure of sentence processing time and experimental control, eye-tracking research indicates it is not a naturalistic way to read (Rayner & Sereno, 1994). In fact, eye movements to previously fixated sentences are commonly interpreted as attempts to resolve comprehension difficulties (Rayner, Juhasz, & Pollatsek, 2005). Without access to previous text, individual differences in comprehension are more difficult to interpret—if given the opportunity to resolve difficulties, would individual differences be reduced or present at all? Not having the opportunity to look back mimics situations in which readers choose not to look back, do not realize the need to look back, or simply do not have the time to look back. However, there are many situations when the time, realization of need, and desire are present and the opportunity to look back could benefit comprehension. For this reason, it is important to know how reading behavior and comprehension vary with access to prior text. Borella, Ghisletta, and Ribaupierre (2011), using structural equation modeling, showed that working memory mediated the relationship between age and narrative text comprehension when the text was absent during test. However, when the text was present during test, age differences were null for

explicit test questions and minimal for inference questions. Although the entire text was available at once during reading, the reduced demand placed on working memory in the text-present-during-test condition may account for the absence of an age effect on text comprehension. Thus, working memory deficits in older adults may challenge their ability to comprehend and remember texts but the challenge may be overcome under some conditions. Allowing access to prior text and measuring individual differences in working memory capacity could indicate whether age-related effects in text recall/ comprehension are a function of this detail.

The Role of Reading Goal

Other work indicates that the type of reading goal provided by the experimenter could moderate the effect of age-related working memory declines on text comprehension. Zwaan (1994) found trade-offs between text-level recall and situation model construction that depended on the reading perspective (literary or news-story) assigned by the experimenter. Specifically, reading with a literary perspective allowed for better memory of surface text but worse memory for the situational model. If adults of different ages naturally take different reading perspectives, this could account for age differences seen. Work by Brébion (2003) showed that when given the option to focus on processing or storage younger and older adults act differently, although achieving comparable comprehension. Older adults focus on sentence processing for accuracy even if it means slowing down and sacrificing storage. Younger adults will sacrifice sentence processing for storage and will reread instead of slowing down. It should be noted that this study also showed that working memory has a weaker relationship to reading comprehension than to age and may not be an especially strong mediator, at least when

not under severe time constraints. As mentioned previously, Stine-Morrow et al. (2006) showed that working memory mediates the effect of age on text comprehension (as measured by sentence processing) but more so when the instructions were to focus on recall versus comprehension. They also found that this was in part due to memory self-efficacy. Presumably, as the criterion for recall increased, older adults' reduced self-efficacy prevented them from allocating resources to meet the recall demand. In summary, different age groups may set different processing goals for themselves which may or may not result in age-related differences in reading comprehension.

The Role of Domain Knowledge

The effect of age-related working memory declines on text comprehension may also depend on domain knowledge. Domain knowledge moderates the effect of age on text comprehension (Dijkstra, 2001; Jeong & Kim, 2009), although not always over and beyond more basic cognitions (Hultsch et al., 1990). It is not clear why domain knowledge reduces age effects but it may enable more time for integration and inference (Miller, Stine-Morrow, Kirkorian, & Conroy, 2004) or more efficient allocation of attention while reading (Soederberg Miller, 2009), which may be critical in individuals with lower working memory capacity. However, although domain knowledge has been shown to be a stronger predictor of listening comprehension than is working memory, it is unable to buffer against its decline (Hambrick & Engle, 2002).

Of course, listening comprehension and written text comprehension are not identical in every way. In younger participants, however, topic familiarity and vocabulary do not necessarily compensate for difficulties in listening (Bransford & Johnson, 1973) or reading (Freebody & Anderson, 1983) comprehension. High topic

familiarity may be most facilitative with low coherence texts, which force compensatory processing in the form of inferences (McNamara, Kintsch, Songer, & Kintsch, 1996). Thus, domain familiarity may differentially benefit older adults to the extent that they experience texts as less coherent than younger adults and attempt to form inferences while reading. This explanation seems to be inconsistent, however, with findings that younger adults perform best with moderate coherence but older adults perform best with high coherence texts (Hess, 1995).

Measurement Method May Matter

The effect of age-related working memory declines on text comprehension may also depend on how comprehension is measured. For instance, Byrd (1985) found larger age differences in summarization quality than in free recall. Specifically, older adults' summaries had fewer details and failed to exclude extraneous information as well as younger adults' summaries. The simultaneous processes of comprehending and organizing information in preparation for summarization, Byrd suggested, may have been too taxing for older adults' working memory. Adams (1991) also found that older adults' free recall and summaries included fewer details than those of younger adults (although she found no difference between summary and recall quality). Although her study did not test this hypothesis directly, Adams attributed the age difference to a working memory deficit. Measuring comprehension with summaries, however, has not always revealed an age-related deficit; Jackson and Kemper (1993) found that summaries written by older adults included more total and central ideas.

Although past work using recognition paradigms has suggested that situation model construction is unimpaired among older adults (Radvansky & Dijkstra, 2007),

these summary studies suggest that may not always be the case. Alternatively, they may suggest that although comprehension may not decline with age, the ability to communicate what one has comprehended does. Only a meta-analysis of age differences in text *memory* (not in text *comprehension*) exists (Johnson, 2003), so it is less clear whether situation model construction really is impaired with age. More studies, as well as a meta-analysis of these studies, would elucidate the conditions under which text comprehension is more and less affected by age. If what is common to situation model construction and summarizing is the extraction of the main idea and important ideas, while editing out unimportant details, soliciting summaries should provide a good measure of the situation model, one which has not yet been examined but has direct real word applications.

Importance of Metacomprehension

Texts read for school, work, and to inform important life decisions require effective metacomprehension—the ability to monitor and control one's text comprehension (Maki & Berry, 1984). In these situations, individuals must be able to evaluate their comprehension of texts and whether it is sufficient to meet their comprehension goal. Furthermore, when necessary, they must be able to effectively implement strategies that will improve their comprehension. Given that many basic cognitive abilities have been shown to decline with age (Verhaeghen & Salthouse, 1997), we might assume that higher order cognitive abilities such as metacomprehension would as well, but empirical tests of this assumption are sparse. The proposed dissertation attempts to fill this gap.

Measuring Metacomprehension

Measures of metacomprehension can be absolute or relative. Absolute measures assess whether one is over or under-confident on a set of texts, whereas relative measures assess whether one can discriminate between texts more and less well understood. Although occasionally metacomprehension is probed with think-aloud methods (Kendeou et al., 2011; Linderholm, Kwon, & Wang, 2011) and other online measures, such as indicating with a keyboard press when experiencing reading difficulty (Dunlosky, Baker, Rawson, & Hertzog, 2006), more often the metamemory tradition is followed with participants asked to make specific judgments (predictions and postdictions) about test performance. Judgments are compared to performance on tests where participants must recognize verbatim or paraphrased sentences from the texts, or answer multiple choice questions to which the answers may have been implicit in the text. To date, there are no reported studies comparing predictions and postdictions to summaries as a way to measure metacomprehension. Determining these correlations and some of the factors involved was the focus of this dissertation.

Metacomprehension in Younger Adulthood

Accuracy of Metacomprehension

Metacomprehension predictions are usually made with above chance accuracy (Maki, Foley, Kajer, Thompson, & Willert, 1990) but accuracy is still quite low (Maki, 1998b). Considerable individual differences, however, have been found, with some persons being highly accurate (Chiang, Therriault, & Franks, 2010). Accuracy is higher and more reliable when there is more than one test question per text (Weaver, 1990), which has not always been the nature of the criterion tests in these studies (Glenberg & Epstein, 1987). Maki (1998a) found that the most accurate text comprehension

predictions were made with immediate judgments followed by immediate test, although Linderholm and Zhao (2008) found that immediate judgments only benefitted those with high working memory capacity. Moreover, some work with delayed summaries suggests that people better assess what they know after a delay (Anderson & Thiede, 2008; Thiede & Anderson, 2003). As with metamemory, postdictions are more accurate than predictions (Pierce & Smith, 2001). These studies informed both the design and the hypotheses in the present study.

The Role of Working Memory Capacity

One cognitive variable that appears to affect metacomprehension ability is working memory capacity. Work by Linderholm and van den Broek (2002) showed that people with lower working memory capacity make fewer metacognitive statements while reading, and they recall less of what they read. However, more recent work has shown that low spans have higher metacognitive accuracy than high spans under high effort (Ikeda & Kitagami, 2012) and easy texts (increased local cohesion (increased local cohesion; Ikeda & Kitagami, 2013).

However, most investigations into the role working memory capacity may play in metacomprehension have employed a rereading paradigm. Immediate re-reading has been shown to improve metacomprehension (Dunlosky & Rawson, 2005; Rawson, Dunlosky, & Theide, 2000) but not invariably so. Presumably, the second reading allows resources used initially for comprehension to be redirected to comprehension monitoring. But, rereading may only benefit those with low working memory capacity (Griffin, Wiley, & Thiede, 2008). Chiang et al. (2010) found that only in a selective rereading condition, which involved the opportunity to look back in text while reading, was

working memory related to metacomprehension. Because selective rereading requires active monitoring of comprehension breakdowns, it may be more demanding than simply reading a passage twice. But metacomprehension accuracy was not significantly different in the rereading and selective rereading conditions, so the role of working memory capacity in metacomprehension is unclear. Walczyk and Taylor (1996) found that working memory accuracy was unrelated to look backs, although working memory latency—time to retrieve target lexical information—was. Looking back in text, or selective rereading, may benefit those with lower working memory capacity to the extent that they are monitoring their comprehension, are aware of comprehension breakdowns, view looking back as a way to ameliorate these breakdowns, and choose to expend the time and effort to exert this control strategy.

Summaries as Metacomprehension Criterion

Although memory of various ideas and details in text can be important, the ability to demonstrate knowledge of what is most important in text, by summarizing, is critical in many areas of life—accurate assessments of summary quality could be essential for determining when rereading may be necessary to reach a comprehension goal. Despite this reality, the ability to judge the quality of one's own summaries has not yet been tested. Summaries have been used in metacomprehension studies as a way of improving the correlation between predictions about test performance and actual test performance (Anderson & Thiede, 2008; Thiede & Anderson, 2003) but not as the metacomprehension criterion itself. When written after a delay, summaries allow participants to judge which texts are not well learned and should therefore be selected for re-study. Franzke, Kintsch, Caccamise, Johnson, and Dooley (2005) showed that 8th-grade student summaries can

improve over cycles of feedback, but they did not assess whether students could predict this improvement or adequately judge the quality of their summaries before or after they were written.

There is only one metacomprehension study in which participants were asked to write and judge summaries (Schommer & Surber, 1986). In this study, however, a multiple choice comprehension test, not the summaries, was the criterion for text comprehension and metacomprehension. The summaries only served as representations of the text under different levels of metacomprehension accuracy. Those who accurately predicted good comprehension had better summaries than those who accurately predicted poor comprehension, capturing more of the superordinate concepts contained in the text. Although those who had an illusion of understanding had sufficient superordinate concepts, their summaries included more distortions. Given that the participants rated their comprehension before they wrote summaries, this study suggests that people might be able to accurately predict the quality of their summaries. This assumes, however, that the cues upon which the judgments are made are the same when predicting multiple-choice test performance as they are when predicting summarizing quality and that these cues are actually diagnostic of future performance. Moreover, Schommer and Suber only administered two passages, so it is still unknown whether with multiple passages people can reliably predict their average summary quality as well as distinguish between summaries of different quality.

Metacomprehension in Older Adulthood

Does Metacomprehension Decline with Age?

The metacomprehension and aging literature is difficult to summarize, in part because methods and criterion measures vary considerably across a relatively small set of studies. Both survey based and experimental analyses of aging and metacomprehension have shown mixed results. The self-report measure, Metacomprehension Scale (Moore, Zabrucky, & Commander, 1997b), has produced different findings in two studies. Lin et al. (2000) showed that older and younger adults report similar metacomprehension but Moore et al. (1997a) showed older adults report lower average values on just two subscales, regulation and achievement. A similar but different questionnaire used in an Italian sample showed older adults reporting worse metacomprehension overall (De Beni et al., 2007). Early experiments operationalized metacomprehension monitoring as the ability to detect inconsistencies in text, finding no age-related declines in this ability (Moore & Zabrucky, 1992; Zabrucky & Moore, 1994; Zabrucky, Moore, & Schultz, 1987) unless working memory load was high (Zabrucky, Moore, & Schultz, 1993). Later experiments asked participants to make predictions about their performance on tests of text comprehension. Some studies showed no age-related differences (Dunlosky et al., 2006; Olin & Zelinski, 1997), whereas others showed age-related deficits (Baker et al., 2010; Miles & Stine-Morrow, 2004). Only one study has looked at age differences in postdictions in metacomprehension (Baker et al., 2010). It found that older adults do not show a deficit in making term specific postdictions, at least when first attempting to retrieve the term definition and then comparing it to the experimenter-provided correct one. In summary, there is insufficient evidence to make robust conclusions about whether older adults are deficient in their ability to monitor comprehension of texts.

It is also ambiguous whether control processes in metacomprehension decline with age. Older adults are less likely to reread when they detect the inconsistencies (Zabrucky & Moore, 1994), suggesting that metacomprehension control is impaired with age. However, older adults may be able but unwilling to reread, for a variety of reasons. When asked to prioritize accuracy over speed, older adults will allocate more time to reading but they do not increase this allocation as much as younger adults do (Stine-Morrow et al., 2006). Other similar work indicates that the way in which older adults allocate their resources, relative to younger adults, varies. Sometimes older adults spend longer at wrap-up (clause and sentence boundaries) to achieve good comprehension (Stine-Morrow et al., 2001; Stine-Morrow et al., 2008), whereas other times they will spend more time at intrasentence boundaries and less time at the ends of sentences (Stine-Morrow, Noh, & Shake, 2010; Stine-Morrow, Shake, et al., 2010; Stine, 1990).

Accounting for Age Differences

Should age-related differences in metacomprehension be real, one intuitive suspect could be working memory. Findings of its decline with age are robust (Verhaeghen & Salthouse, 1997) and at least some aspects of metacognition have been shown to be resource demanding (Souchay & Isingrini, 2004; Zabrucky & Moore, 1994). Older adults may have more trouble holding in mind information presented earlier in text in order to integrate it with later text. In point of fact, Zabrucky, Moore, and Schultz (1993) found that older adults had more trouble evaluating whether a target sentence was inconsistent with a cue-sentence when they were separated by four filler sentences, as opposed to one. Alternatively, older adults may have more trouble concurrently reading texts and monitoring for comprehension. For example, Miles and Stine-Morrow (2004)

found a group of older adults with lower working memory capacity were less good at predicting how much they would learn. Furthermore, the work by Stine-Morrow and colleagues on age differences in wrap-up behavior, summarized earlier, suggests that older adults attempt to compensate for working memory declines when possible—e.g., spending extra time at intrasentence boundaries may reduce the load at the ends of sentences (Stine-Morrow, Shake, et al., 2010). However, reduced working memory capacity does not always account for poorer metacomprehension. In Olin and Zelinski's (1997) sample, younger adults had slightly higher working memory capacity but working memory was unrelated to any of the dependent measures. Moreover, Linderholm and Zhao (2008) found that younger adults with lower working memory capacity had *better* absolute metacomprehension and that those with higher working memory capacity were best when they had to process more deeply. Those with lower working memory capacity benefit in metacomprehension, the authors postulated, because their reading was more effortful and because they have more activated concepts (due to less inhibition). Although it is not clear whether there are age differences in metacomprehension or what may account for them, young adult studies, summarized earlier, as well as some aging studies, suggest that working memory deficits in older adults could account for metacomprehension deficits under some circumstances.

Current Study

The current study sought to build on previous literature by examining whether there are age-related differences in metacomprehension as measured by summaries and summary judgment. Examining metacomprehension with summaries as the criterion is worthwhile given that outside of lab experiments and school settings comprehension is

often demonstrated in summary form. Both recognition tests and summaries can show what a person remembers and understands, but summaries measure situation models in a way that traditional comprehension tests do not. Summaries force the summarizer to decide what is important without any of the visual cues provided by recognition and multiple choice tests. However, because summaries can be limited by one's ability or desire to demonstrate what they understand, multiple choice tests were also used as a check on text comprehension and as a link to earlier metacomprehension literature.

Although summaries have been used to measure age-related differences in text comprehension (Adams, 1991; Byrd, 1985; Jackson & Kemper, 1993), they have not been used to measure metacomprehension in older adults and in only limited ways in younger adults (Schommer & Surber, 1986). Research suggests that older adults are capable of constructing an adequate situation model (Radvansky & Dijkstra, 2007), but it is not known whether they can adequately judge this ability.

The effect that reading goals have on age-related differences in comprehension and metacomprehension is also essentially unknown. Although older adults adjust less to experimenter instruction to focus on speed or accuracy (Stine-Morrow et al., 2006), it is not known whether older adults adjust less than younger adults for other reading goals. Younger adult studies have shown that they read and remember texts differently for given different goals (Dudukovic et al., 2004; Kendeou et al., 2011; Linderholm et al., 2008; Linderholm & van den Broek, 2002; Linderholm & Wilde, 2010; Narvaez et al., 1999; Van den Broek et al., 2001), but it is yet to be determined whether the reading/summarizing goal has the same effect on older adults. Although younger adults tend to rate their comprehension higher when reading with a study goal (Linderholm et

al., 2008; Linderholm & Wilde, 2010) this is not necessarily an accurate judgment (Linderholm & Wilde, 2010). It remains to be seen whether reading goals affect younger adult summary *judgment* in the same way and whether reading goals affect older adult summary judgment at all.

Metacomprehension accuracy and age differences therein may be dependent on working memory capacity and rereading behavior. Given that older adults have poorer working memory, age-related differences in text comprehension and metacomprehension may be expected to the extent that they tax this capacity. However, selective rereading may be able to compensate for lower working memory capacity, but most studies have prevented access to prior text so it is unknown how rereading behaviors may interact with working memory capacity to influence metacomprehension. Access to prior text may benefit those with lower working memory capacity but it may also allow those with higher working memory to review and further outperform those with lower capacity. Rereading (reading the full text twice) has been shown to improve metacomprehension in younger adults (Rawson et al., 2000) but only one study (Chiang et al., 2010) has looked at the effect of selective rereading on metacomprehension. Selective rereading has never been examined in older adults, a population for whom access to prior text could matter, especially given their working memory deficits. For this reason, I provided access to prior text in the present study in order to test the effects of selective rereading on comprehension and metacomprehension directly as well as a moderator of the potentially mediating effect of working memory capacity. The interactive effects of age, reading goal, selective rereading, and working memory capacity on metacomprehension ability were also analyzed.

In addition to examining the effect of reading goal and access to prior text on summarization judgments, the relationship between summary judgment ability and both self-reported metacomprehension and reading self-efficacy was measured. These self-report measures were administered before and after the main experiment so that the change in these constructs due to summary writing and judgment could be assessed. Working memory was measured with the reading span task in order to determine whether it can account for any individual differences in summary judgment that may be found. Older adults show declines in working memory on average (Salthouse, 1994), but support for a resource deficit account of age differences in text comprehension is mixed (Hartley, 1993b; Hultsch et al., 1998; Van der Linden et al., 1999) and may be the consequence of subtle differences among tasks. I anticipated that concurrently reading, organizing what is read in preparation for summarizing, and monitoring text comprehension could be more resource demanding than tasks previously used in this area of study and thus a measure of working memory would be necessary to interpret results. This hypothesis would be supported if working memory capacity is found to account for a significant proportion of variance in summary judgment ability. At the study end, participants also completed a vocabulary test, a survey of their reading habits, and ratings of text difficulty, text topic familiarity, and text interest level, which can be examined as possible covariates in the future.

Research Questions

1) Can adults accurately judge the quality of their text summaries before and after summarizing?
2) Are there age-related differences in these abilities?

3) Do summarizing goals (summarize for a professor/boss vs. acquaintance/stranger) influence these abilities?
4) Does access to prior text predict these abilities?
5) Does age interact with either reading goal or access to prior text to influence these abilities?
6) Does working memory capacity account for between person variation in summary quality judgment?
    a. If so, does the relationship between working memory capacity and summary quality judgment depend on reading goal or access to prior text?
    b. If so, do age and working memory capacity have an interactive effect on summary quality judgment?

Hypotheses

Ability to Judge Summaries

I expected participants as a whole to be poor at judging their summary quality, although more accurately postdict than predict. I expected participants to have above chance accuracy in discriminating between texts more and less well summarized, with judgments of details to be more accurate than those of main ideas, as would be suggested by work by Weaver and Bryant (1995) that showed that performance on detail questions over expository texts are more accurately predicted than thematic questions.

Main Effects of Reading Goal

To the extent that texts read with the acquaintance/stranger goal might be processed more shallowly, summary predictions for acquaintance/stranger texts were expected to be less accurate than those for the professor/boss goal given that shallower

processing may be related to poorer metacomprehension (Linderholm & Zhao, 2008; Schommer & Surber, 1986).

Role of Access to Prior Text (Rereading)

Access to prior text was expected to allow participants to reinstate information earlier in text that has been lost from working memory so that it could be used to build a coherent situation model, which would benefit metacomprehension accuracy. However, if those who struggle to remember and/or comprehend are unaware of their difficulty and do not take advantage of the opportunity to go back or compensate in some other way, access to prior text behavior was expected to not predict text comprehension or metacomprehension. There would also be no relationship between this behavior and the dependent variables if the task could be accomplished successfully by most people without having to go back. This was expected to occur if the texts were not difficult to read, although the optimal text difficulty is different for young and older adults so there may be an interactive effect with age. It also may not occur if participants slow their reading for each sentence or read each sentence more than once before proceeding, precluding the need to go back to previous sentences.

Main Effects of Age

I predicted that older adults would produce similar main ideas to younger adults, but fewer details, and that their judgments would reflect this difference. I predicted that older adults would be well calibrated on main ideas (significant text-level relations of ratings and summary quality) but overconfident about details; in contrast, younger adults would be overconfident on both main ideas and details.

Age x Reading Goal Interactions

I predicted that age differences would be larger for the professor/boss goal than for the acquaintance/stranger goal, in part because older adults adjust less to experiment goal instruction (Stine-Morrow et al., 2006). The standard for the professor/boss goal was expected to be higher than that for an acquaintance/stranger and the increase in resource demands to achieve this standard may exceed the average capabilities of older adults. Thus, older adults may value and attempt to achieve this higher standard for comprehension but fail to do so. A possible interaction of summarize condition with age (because of working memory deficits) with underconfidence for older adults in the professor/boss condition but overconfidence in the acquaintance/stranger condition (Linderholm & Zhao, 2008) was also anticipated.

Age x Access to Prior Text Interactions

Access to prior text allows a reader to refresh forgotten information and resolve confusion. Older adults generally have lower working memory capacity and access to prior text was predicted to allow them to overcome this deficit and achieve comparable performance to younger adults. However, older adults do not always selectively reread when they do not understand (Zabrucky & Moore, 1994) so they may not use this feature to their benefit. If only younger adults take advantage of this feature, they may use it to achieve superior performance to older adults. If younger adults do not need to access prior text to reinstate it and older adults do not take advantage of it, or if neither group needs the opportunity, I expected to fail to find a relationship between this behavior and the dependent variables.

Relationship between WMC and comprehension/metacomprehension

To the extent that older adults differ from younger adults on either the main task, in working memory capacity, or both, I expected to determine the role that working memory capacity plays in age differences or the lack thereof. If the benefit of working memory capacity is to allow one to simultaneously attend to comprehension and monitoring of comprehension I expected to find that increased working memory capacity is associated with better comprehension and metacomprehension measures. If the relationship between working memory capacity and test comprehension or metacomprehension is dependent on task difficulty (Ikeda & Kitagami, 2012) I expected to see its effects moderated by either reading goal or access to prior text.

CHAPTER 2

METHODS

Design

The ability to predict and postdict summary quality was measured in a 2 (Age: Younger, Older) X 2 (Summary goal: professor/boss, acquaintance/stranger) between subjects design.

Participants

Participants in this study were 141 younger adults (ages 18-30, *M* = 19.40, *SE* = .14), 54 women) recruited from the Georgia Tech subject pool, compensated with one course credit per hour, and 138 older adults (ages 60-80, *M* = 69.38, *SE* = .48, 84 women) recruited from the Adult Cognition and Development Lab's participant database, compensated with $10 per hour. Younger adults reported an average of 14.26 (*SE* = .99) years of education. Older adults reported an average of 17.36 (*SE* = 1.95) years of education.

Materials and Procedure

Six expository passages from Experiment 3 of Rawson and Dunlosky (2002) were used. These passages were adapted from the Scholastic Aptitude Test (Board, 1997), written at Flesch-Kincaid grade-level of 9.8-12.0 (*M* = 11.6), with a Flesch readability score of 22.1-62.2 (*M* = 44.3). They are 370 words in length on average. Titles of these passages include: Television Newscast, Precision of Science, Women in the Workplace, Zoo Habitats, American Indians, and Real vs Fake Art (see Appendix A for sample passage). Passages were presented with 18-point bold Courier New font on personal computers.

Participants began by providing informed consent, completing the PDS, and completing the first MCS and reading self-efficacy questionnaires. Individuals were then presented with six passages with either instructions to summarize for a professor (younger adults only) or boss (older adults only) or instructions to summarize for an acquaintance/stranger (younger and older adults). Passage presentation order was randomized and each passage was presented one sentence at a time. Participants were told to press the space bar anytime they experienced processing difficulty and that they could move backwards in the text by pressing the B button and forward by pressing the N button, as often as they liked. Before each new passage was presented, they were encouraged to take advantage of this feature. After reading each passage, participants made predictions about their ability to summarize that passage. Before summarizing and postdictions, participants completed the Reading Span test. In addition to serving as a measure of working memory capacity, this test was given to clear working memory of the most recent text. Then, Audacity (http://audacity.sourceforge.net/) was used to record oral summaries, prompted by the title of each story on a slide in E-prime (Schneider, Eschman, & Zuccolotto, 2002). Summarizing was done orally to minimize writing fatigue, especially for older adults who may have arthritis, and to avoid excessive experiment length (pilot testing showed that written summarizing took twice as long as oral summarizing). Although oral summaries may include more false starts, errors, and additions (Harker et al., 1982; Vieiro & García-Madruga, 1997), the modality of text recall does not affect the size of the age difference (Bryan & Luszcz, 1999; Johnson, 2003). After reporting each summary, they pressed the Enter key and were asked to make postdictions about the quality of that summary with a keypress in Eprime. Next

they answered eight multiple choice questions about each text they read and made a judgment about how many out of each set of eight they thought they answered correctly. Finally, participants completed the vocabulary test, second instances of the MCS and reading self-efficacy measures, reading habits questionnaire, and text ratings questionnaire.

Summarizing Instructions & Schema Induction

Participants were told that their summaries should contain only the main ideas and important details and should not be a list of everything they remember from the text. They were also told that their summaries would be given points for main ideas, important details, the presence of a topic sentence, and efficiency (capturing the main ideas and important details succinctly) but penalized for the inclusion of unimportant details or redundant information. Those who read and summarized with the professor/boss goal were asked:

> "Please take a minute or so to describe a time you had to read to prepare for an oral summary for a professor or boss. If you have never given an oral summary, describe a time you had to read to prepare for an essay test."

After doing this they were asked:

> "Please describe what you think would make a good summary when summarizing for a professor/boss."

After this schema induction, they were reminded to read the texts as if they were planning to give an oral summary for a professor (YA) or to their boss (OA).

Those who read and summarized texts with the acquaintance/stranger goal were asked:

"Please take a minute or so to describe a time you told an acquaintance/stranger about something you read, something that was based on facts."

After doing this they were asked:

"Please describe what you think would make a good summary when summarizing for an acquaintance/stranger."

After this schema induction, they were reminded to read the texts as if they were planning to tell an acquaintance/stranger about them.

Metacomprehension Judgments

After reading each text, participants were asked to predict their summary quality by answering the following questions with a number keypress:

1. How well do you think you will be able to summarize this passage? (Likert response scale 1-7)
2. What percentage of the main ideas in this passage do you think your summary will include?
3. What percentage of the important details in this passage do you think your summary will include?
4. How many questions out of eight do you think you will answer correctly about this passage?
5. How well do you think you could generate the theme to the passage you just read? (Likert response scale 1-7)

[The participants then orally stated what they believe to be the theme of the passage]

6. How well do you think the theme you just generated captures the actual theme of the passage you just read? (Likert response scale 1-7)

After orally summarizing each text, participants were asked to judge the quality of their summary by answering the following questions:

1. How well do you think you summarized the passage? (Likert response scale 1-7)
2. What percentage of the main ideas from the passage do you think you included in your summary?
3. What percentage of the important details do you think you included in your summary?

Comprehension Test

Comprehension was measured with summaries and a multiple forced choice test taken from Rawson and Dunlosky (2002), the same origin of the texts borrowed for this study. The test consisted of eight questions per text, four of which could be answered by information explicitly stated in the text, and four which require inferences (See Appendix A). After answering the questions for each text, participants were asked, "How many questions out of eight do you think you answered correctly about this passage?" Metacomprehension judgments were scaled to the summaries and multiple choice test performance. Of primary interest was participants' ability to judge their summary quality, but the multiple choice test was used as an additional check for comprehension and as a link to previous literature.

Survey Measures

Demographic Survey

The short form of the Participant Demographic Survey (PDS) typically used in our lab was administered.

Metacomprehension Scale

The Metacomprehension Scale (MCS; Moore et al., 1997b) is a measure of self-perceived metacomprehension ability. It is composed of seven subscales (regulation, strategy, task, capacity, anxiety, achievement, and locus) which have been confirmed using principal component analysis, determined to hold both convergent and discriminant validity, and account for 19% of comprehension variance. It was administered to allow for comparison of perceived and actual metacomprehension abilities, which has not yet been reported in the literature. The MCS was administered at the beginning and end of the experiment in order to assess whether perceived metacomprehension abilities are altered by task experience.

Reading Self Efficacy

This survey was borrowed from Prat-Sala and Redford (2010). Reading self-efficacy has been shown to be related to essay writing in college students (Prat-Sala & Redford, 2012). This self-report measure was administered to assess whether reading self-efficacy is also related to oral summaries and whether differentially so in younger and older adults. It was also administered to see if the relationship between reading self-efficacy and metacomprehension is similar to that between memory self-efficacy and metamemory, including whether it might account for age-related differences therein. It

was administered at the beginning and end of the experiment to assess whether reading self-efficacy can be altered by task experience.

Reading Habits

A survey adapted from the Reading Habits Self-Report (Acheson, Wells, & MacDonal, 2008) was administered to query the reading and summarizing habits of participants. It consisted of questions about reading purposes, such as 'for school or work' and 'for pleasure', which required participants to respond from among six choices ranging from 'Never' to 'More than 3 hours per day'. It also consisted of questions about how often participants read or summarized various reading materials, to which responses were to be made on a 0 to 7+ hour Likert scale. Reading habits can be examined as a potential covariate, which also may be age-related.

Text Ratings

For each text, participants were asked to indicate, on three separate Likert scales (1-7), how interesting, familiar, and difficult each text was that they read. Self-reported interest, familiarity, and difficulty of texts may be potential covariates of interest that could assist in the interpretation of results.

Other Cognitive Measures

Vocabulary Test

The Shipley Institute of Living Scale – Revised (Zachary, 1986) is a 40-item vocabulary test. Participants have 10 minutes to decide which of four possible synonyms most closely matches the meaning of the cue-word. Verbal ability was measured as a potential mediator of age-related differences in metacomprehension ability and as a link to the aging and text comprehension literature.

Reading Span

A computerized version of the Reading Span (Redick et al., 2012; Unsworth, Heitz, Schrock, & Engle, 2005), a measure of working memory capacity, was administered using E-prime software (Schneider et al., 2002). This task takes about 12 minutes to complete.

Summary Scoring

For the purpose of assessing participant summaries of read passages, a rubric was created outlining the theme, main ideas, and important details of each passage. A team of four, the author and three undergraduate students, developed the rubric. The author developed the first draft of the rubric, which was then modified by the group. Through a series of cycles, we attempted to independently score a random selection of outputs from participants and used discussion of discrepancies to modify the rubric and develop rules for applying the rubric (See Appendices B and C).

Each participant generated summary was scored for the presence of each main idea and important detail identified in the rubric. The number of each was divided by the total possible and converted to a percent. In scoring the presence of main ideas and important details, the standard was gist consistency, as the goal was to determine whether comprehension, not verbatim recall, was demonstrated.

Themes and overall summarization were given a rating on a 7-pt scale. Initially we planned to give each summary a composite score such that points would be given for the presence of a topic sentence and efficiency and points would be deducted for the presence of redundant or unimportant information. Given the challenges in establishing inter-rater reliability, including the extensive time it took to achieve reliability on the

crucial measures, we decided to modify the scoring of these peripheral characteristics. Instead of a topic sentence, per se, we looked for overall organization and coherence, and we considered redundant and unimportant information both as elements of efficiency. Scoring of the overall summary quality took into account the number of main ideas and important details as well as the organization, coherence, efficiency, and presence of inaccuracies.

Inter-rater reliability was calculated weekly during rubric development. In the educational psychology literature, 50% agreement is acceptable by some when decisions are made about group differences and not about an individual's fate (e.g. determining college entrance; Cherry & Meyer, 1993). Also in the educational psychology literature, a one-point disagreement in scores is treated by taking the average of two scores and does not deflate the reliability—the observed score is a combination of the true score and error so you cannot differentiate between a situation in which one rater gives a score of two and another rater gives a score of four versus a situation in which both raters give a score of three because how the error is distributed around the scores is unknown (Johnson, Penny, & Gordon, 2001). Only one undergraduate rater reached a level of inter-rater reliability with the author to continue to rate materials. This rater and the author reached a rate of 74.5% exact agreement and 97% 1-pt difference agreement for a consecutive 10% of subjects. At this point the remaining data was divided for rating purposes.

The intraclass correlation was also calculated as a measure of inter-rater reliability. The ICC is a ratio of between-subject variability to total variability (which includes rater variability). As the ratio approaches 1, much more variability is due to subjects than raters. The ICC was above 75% for all dependent measures.

CHAPTER 3

RESULTS

Analyses

SAS Proc Mixed (Littell, Milliken, Stroup, & Wolfinger, 2000) was used for data analysis. For metacognitive research questions, traditional analyses were performed first, as a point of comparison to previous literature. As a measure of relative accuracy, Stuart's tau-c (Stuart, 1977) was used instead of the traditionally used Goodman-Kruskal gamma correlation (Nelson, 1984). Like gamma, tau-c is a within-person correlation that provides a relative measure of the correspondence between objective scores and subjective scores and therefore assessed whether participants can discriminate among varying levels of passage comprehension. Tau-c was used instead of gamma because, with only 6 passages in this study, there is limited opportunity for scores to vary and a high percentage of gammas can be at the marginal means resulting from the throwing out of ties. Tau-c has the advantage of correcting for ties and is a more conservative estimate of the ordinal relationship of scores and judgments. Bias, an absolute measure of summary quality, was also calculated. For each person, the mean difference (across summaries) between their actual and judged performance was determined, separately for predictions and postdictions. Bias indicates how under- or over-confident participants are, on average, in their summarization abilities.

Multi-level modeling (MLM) was also used whenever passage level data was of interest. A critical issue for assessing summarization quality is to link within-person, between-passage variation in metacognitive judgments and comprehension performance, while also enabling cross-level effects of person-level variables, such as age, on these relationships. Traditionally, metacognitive researchers have used a two-stage process,

using gamma correlations to generate ordinal associations of within-person relations of judgments and performance for each individual, and then using standard GLM procedures to evaluate individual differences in these correlations. This approach is deficient, statistically speaking, in several respects, including computation of incorrect standard errors on cross-level regression coefficients and the sensitivity of gamma to skewed marginal distributions. For this reason, MLM was used to examine within- and between person variability in performance, building in specific contrasts to test interactions with manipulated context. This technique is highly appropriate for analyzing experimental effects while avoiding aggregation bias (Hoffman & Rovine, 2007), and it is being increasingly used to analyze experiments that study aging and cognitive mechanisms in reading and metacognition (e.g., Hines, Touron, & Hertzog, 2009; Payne et al., 2012).

Centering predictors is a part of multi-level modeling that is done to address multicollinearity and create more interpretable results, among other reasons. There are different ways to center predictors which allow for addressing different research questions. For this dissertation, I chose to center within age group so that I could focus on interactions of level-1 and level-2 (age) variables, while ignoring the issue of mediators for between-age groups differences. With an extreme age groups design, centering at the grand mean for variables such as working memory would not have been ideal as grand mean centering assumes that group means are uncorrelated with the predictors (Algina & Swaminathan, 2011)

Vocabulary, Working Memory, and Rereading

As is usually found, older adults had better vocabulary [$F(1,275) = 46.45$, $p < .001$] but lower working memory capacity [$F(1,273) = 67.36$, $p < .0001$] than did younger

adults. Older adults also reread more often than younger adults ($F(1,279) = 4.34$, $p = .04$ (See Table 1). There was no effect of reading goal on vocabulary, working memory, or rereading.

Passage Ratings and Reading Habits

There were significant age differences in ratings of passage difficulty, familiarity, and interest. For difficulty ratings, only an age by passage interaction was significant [$F(5,1375) = 612.80$, $p < .0001$]. Compared to older adults, younger adults reported greater difficulty with all passages except the science passage. For familiarity and interest ratings, there were main effects of age qualified by significant age by passage interactions. Older adults reported more familiarity with [$F(5,1374) = 26.16$, $p < .0001$] and greater interest in [($F(5, 1374) = 15.95$, $p < .0001$] all passages except the science one, compared to younger adults. There were also significant individual differences in these ratings (difficulty variance = 1.02, $SE = .04$, $p < .0001$; familiarity variance = 0.77, $SE = .09$, $p < .0001$; interest variance = 0.90, $SE = .10$, $p < .0001$). Older adults also reported more reading and more summarizing per week in their daily lives (See Table 1) than did younger adults. Reading goal had no effect on passage ratings or reported reading and summarizing habits.

Table 1. Participant vocabulary, working memory, rereading frequency, passage ratings, and weekly reading and summarizing behavior.

	Young	Old	*d*
Vocabulary	31.70 (0.30)	34.61 (0.30)*	0.82
Working Memory (Rspan)	59.80 (1.17)	46.06 (1.19)*	-0.99
Rereading	1.59 (0.17)	2.58 (0.18)*	0.48
Passage Interest	4.14 (0.05)	5.14 (0.05)*	1.69
Passage Difficulty	2.81 (0.05)	2.59 (0.05)*	-0.37
Passage Familiarity	3.90 (0.05)	4.19 (0.05)*	0.49
Weekly Reading Frequency (hrs)	2.23 (0.09)	2.90 (0.10)*	0.60
Weekly Summarizing Frequency (hrs)	0.85 (0.08)	1.26 (0.09)*	0.41

Note: Means with standard errors in parentheses. * $p < .05$; d = Cohen's d effect size

Self-Reported Metacomprehension (MCS) and Reading Self-Efficacy (RSE)

MCS scores changed significantly from pre to post on most subscales, with participants reporting more anxiety, more value of comprehension, less understanding of comprehension processes, and less confidence in their comprehension ability at test. RSE scores also changed significantly after task experience, with post-RSE lower than pre-RSE (See Table 2). Hence, it seemed that experience with summarizing passages as it was required in this study reduced self-confidence in reading and comprehension ability.

There were significant age differences on most subscales of the MCS for both the pre- and post-administration of the survey (See Table 2). On both pre and post MCS, younger adults were more anxious, valued comprehension less, were less confident in their comprehension abilities, felt they had more internal locus of control, and reported regulating their comprehension less than older adults. On the post-MCS only, younger

adults also reported using fewer reading strategies. Also on the post-MCS, age interacted with reading goal such that understanding of comprehension processes was lower for younger adults in the professor condition but lower for older adults in the acquaintance condition [$F(1,273) = 7.35$, $p < .01$]. There were no differences due to reading goal on any pre-MCS measures.

There was no effect of age on pre- or post-RSE (See Table 2) or an effect of reading goal on pre-RSE [$F(1,270) = .46$, $p = .50$] or post-RSE [$F(1,272) = .08$, $p = .78$].

Table 2. Metacomprehension Scale and reading self-efficacy ratings as a function of age and time.

	Pre			Post		
	Young (n = 141)	Old (n = 138)		Young (n = 141)	Old (n = 138)	
MCS Anxiety	2.54 (.09)	2.23 (.09)	*Age d* = 0.29*	3.05 (.10)	2.75(.10)	*Age d* = 0.25*
	2.39 (.06)			2.90 (.07)		
	t (272) = - 10.58, *change d* = 1.28**					
MCS Achievement	4.45(.04)	4.77 (.04)	*Age d* = -0.68*	4.56 (.04)	4.83(.04)	*Age d* = -0.57*
	4.61 (.03)			4.69(.03)		
	t (273) = - 4.59, *change d* = 0.56**					
MCS Strategy	3.01 (.07)	3.07 (.07)	*Age d* = -0.07	2.96(.08)	3.25 (.08)	*Age d* = -0.31*
	3.05 (.05)			3.11 (.05)		
	t (274) = - 1.56, *change d* = 0.19					
MCS Capacity	4.24 (.05)	4.47 (.06)	*Age d* = -0.35*	3.90 (.06)	4.12 (.06)	*Age d* = -0.31*
	4.36 (.04)			4.01(.04)		
	t (272) = 11.42, *change d* = -1.38**					
MCS Task	1.65 (.06)	1.70 (.06)	*Age d* = -0.07	1.82 (.06)	1.78 (.06)	*Age d* = 0.05
	1.68 (.04)			1.80 (.05)		
	t (273) = - 2.91, *change d* = 0.35*					
MCS Locus	3.37(.05)	3.59 (.05)	*Age d* = 0.37*	3.45 (.05)	3.60 (.05)	*Age d* = -0.25*
	3.47 (.03)			3.53 (.03)		
	t (271) = - 1.88, *change d* = 0.23					
MCS Regulation	3.89 (.05)	4.20 (.05)	*Age d* = -0.53*	3.82 (.06)	4.18 (.06)	*Age d* = -0.51*
	4.04 (.04)			4.00 (.04)		
	t (272) = 1.28, *change d* = - 0.16					
RSE	5.32 (.06)	5.33(.07)	*Age d* = -0.01	5.02 (.07)	5.07(.07)	*Age d* = -0.06
	5.33 (.05)			5.04 (.05)		
	t (272) = 8.68, *change d* = -1.05**					

Note: Means with standard errors in parentheses; "*Age d*" refers to Cohen's effect size for the difference in age groups. "*change d*" refers to Cohen's effect size for the difference between pre and post ratings; * $p < .05$; ** $p < .001$.

Comprehension scores were correlated with RSE and one MCS subscale, the subscale that directly assesses participants' confidence in their comprehension abilities (MCS Capacity). This was true at study and test, with one exception--MCS1 Capacity was not significantly correlated with percentage of important details generated. The reduction in confidence after task experience resulted in higher correlations between task performance and perception of one's overall reading comprehension abilities with one exception--the correlation of theme scores with RSE at test was not higher than it was at study (see Table 3). Importantly, MLM indicated that only RSE and MCS Capacity after test significantly predicted comprehension scores, with no age differences therein (See Table 4)

Table 3. Correlations of comprehension scores with metacomprehension scale ratings and reading self-efficacy.

	MCS1cap	MCS2cap	RSE1	RSE2
Theme Scores	*r*(275) = .20 **	*r* (277) - .23 **	*r* (274) = .23 **	*r* (276) = .23 **
Percentage Main Ideas	*r* (275) = .12 *	*r* (277) = .22 **	*r* (274) = .17 *	*r* (276) = .27 **
Percentage Important Details	*r* (275) = .11	*r* (277) = .16 *	*r* (274) = .14 *	*r* (276) = 21 **
Overall Summary Quality	*r* (275) = .17 *	*r* (277) = .24 **	*r* (274) = .20 **	*r* (276) = .26 **
Multiple Choice Scores	*r* (275) = .23 **	*r* (277) = .29 **	*r* (274) = .24 **	*r* (276) = .33 **

Note: MCS1cap and MCS2cap refer to MCS Capacity at study and test, respectively. Mean correlations with degrees of freedom in parentheses. * $p < .05$; ** $p < .001$

Table 4. *F* tests for fixed effects of MCS Capacity, RSE, and age on comprehension.

	Themes		Main Ideas		Important Details	
	F	*df*	*F*	*df*	*F*	*df*
MCS1cap	0.17	1,1359	1.20	1,1360	0.01	1,1360
MCS2cap	7.45*	1,1359	16.07*	1,1360	5.57*	1,1360
MCS1cap x Age	0.02	1,1359	2.26	1,1360	0.98	1,1360
MCS2cap x Age	0.01	1,1359	1.20	1,1360	0.18	1,1360
RSE1	1.29	1,1359	2.70	1,1360	0.88	1,1360
RSE2	3.31	1,1359	22.49**	1,1360	11.44**	1,1360
RSE1 x Age	0.94	1,1359	0.07	1,1360	0.30	1,1360
RSE2 x Age	1.25	1,1359	0.01	1,1360	1.05	1,1360

	Overall Summary		Multiple Choice	
	F	*df*	*F*	*df*
MCS1cap	0.33	1,1354	0.71	1,1365
MCS2cap	15.16**	1,1354	13.84*	1,1365
MCS1cap x Age	1.81	1,1354	1.36	1,1365
MCS2cap x Age	0.72	1,1354	0.12	1,1365
RSE1	0.91	1,1354	0.09	1,1365
RSE2	17.98**	1,1354	18.79**	1,1365
RSE1 x Age	0.17	1,1354	0.08	1,1365
RSE2 x Age	0.68	1,1354	0.13	1,1365

Note: MCS1cap and MCS2cap refer to MCS Capacity at study and test, respectively. * $p < .05$. ** $p < .001$.

In summary, the average metacomprehension that younger adults reported (but not reading self-efficacy), was lower than that reported by older adults, participants' lowered their ratings after task experience, and these lower ratings appeared to a more accurate perception of their actual comprehension for both age groups.

Text Comprehension

See Table 5 for mean comprehension scores by age. See Tables 6-15 for MLM significance tests. Fixed effects and random effects are reported in separate tables. Tables for all dependent measures include 3 models. Model 1 includes passage, age, and reading goal as predictors; Model 2 adds passage-level and person-level rereading; Model 3 adds working memory capacity. All models included subject as a random effect. Tables of random effects include an unconditional model (person intercept and residual variance), residual, intercept (random effect, or individual differences), Overall R^2 (variance accounted for by fixed and random effects), and Fixed-effect R^2. Random variances on effects indicate individual differences in those effects. There were significant individual differences in all comprehension measures (See Intercept row of Random Variance Tables).

Passage had interactive effects on all measures, although the pattern of these effects varied across dependent variables. The age by passage interactive effects suggests that the precise nature of the passages need to be considered when conducting research on age differences in passage comprehension. However, for tractability, passage effects will not be described here.

Younger adults performed better than older adults on all measures of comprehension except themes, for which there was no age difference. There was no main

effect of reading goal on any measure of comprehension, but, in Model 3, it had a two-way interactive effect with passage-level rereading on multiple choice scores, as well as a three-way interactive effect with passage-level rereading and age. Although multiple choice scores tended to increase with more rereading across passages, the effect was reduced for those in the professor/boss condition, but not as much for younger adults. The only other effects of rereading were not significant when controlling on working memory (theme scores increased with more within-person rereading and important details increased with more between-person rereading). Higher working memory capacity was associated with higher comprehension scores on all dependent measures.

Table 5. Comprehension scores by age.

	Young	Old	
Theme	5.25 (.07)	5.10 (.07)	d = 0.18
Main Ideas	55.36 (1.21)	49.63 (1.22)	d = 0.40 *
Important Details	30.39 (1.18)	22.99 (1.19)	d = 0.53*
Overall Summary	4.01 (.06)	3.67 (.06)	d = 0.48 *
Multiple Choice	4.87 (.10)	4.11 (.10)	d = 0.64*

Note: Means with standard errors in parentheses. d = Cohen's d effect size; * $p < .05$.

Table 6. *F* tests for fixed effects of passage, age, reading goal, rereading, and working memory on theme scores.

	Model 1		Model 2		Model 3	
	F	*df*	*F*	*df*	*F*	*df*
Passage	26.23*	5,1390	24.50*	5,1398	24.23*	5,1384
Age	2.25	1,279	1.91	1,287	2.01	1,281
Passage x Age	5.14*	5,1390	4.50*	5,1397	4.14*	5,1383
Reading Goal	1.27	1,279	1.68	1,286	2.41	1,281
Age X RG	3.34	1,279	3.40	1,282	4.29*	1,275
RRpassage			4.04*	1,1479	2.76	1,1467
RRperson			0.17	1,315	1.20	1,298
Rspan					6.05*	1,277

* $p < .05$

Table 7. Random variance components for mixed models. Theme scores.

	Model 1		Model 2		Model 3	
	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	1.97*	.07	1.97*	.07	1.97*	.07
Residual	1.31*	.05	1.29*	.05	1.25*	.05
Intercept	0.51*	.06	0.51*	.06	0.47*	.06
Overall R^2	.34	-	0.35	-	.37	-
Fixed-effect R^2	.08	-	0.09	-	.13	-

Var = variance; * $p < .05$

Table 8. *F* tests for fixed effects of passage, age, reading goal, rereading, and working memory on percentage of main ideas.

	Model 1		Model 2		Model 3	
	F	*df*	*F*	*df*	*F*	*df*
Passage	133.96*	5,1391	128.05*	5,1402	131.22*	5,1394
Age	11.17*	1,279	9.23*	1,289	10.18*	1,289
Passage x Age	5.55*	5,1391	4.85*	5,1400	5.07*	5,1392
Reading Goal	0.18	1,279	0.51	1,288	0.57	1,289
Age X RG	0.14	1,279	0.14	1,283	0.02	1,282
RRpassage			0.05	1,1500	0.22	1,1495
RRperson			0.03	1,328	0.26	1,316
Rspan					16.70*	1,284

* $p < .05$

Table 9. Random variance components for mixed models. Percentage Main Ideas.

	Model 1		Model 2		Model 3	
	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	854.01*	29.56	854.01*	29.56	854.01*	29.56
Residual	508.81*	19.30	491.99*	18.66	479.66*	18.26
Intercept	119.09*	17.59	119.73*	17.38	102.18*	15.80
Overall R^2	.40	-	.42	-	.86	-
Fixed-effect R^2	.26	-	.28	-	.74	-

Var = variance; * $p < .05$

Table 10. *F* tests for fixed effects of passage, age, reading goal, rereading, and working memory on percentage of important details.

	Model 1		Model 2		Model 3	
	F	*df*	*F*	*df*	*F*	*df*
Passage	93.49*	5,1390	93.28*	5,1397	93.41*	5,1383
Age	19.45*	1,279	18.47*	1,285	18.14*	1,278
Passage x Age	4.70*	5,1390	5.57*	5,1396	5.83*	5,1382
Reading Goal	0.65	1,279	0.48	1,284	0.32	1,278
Age X RG	0.28	1,279	0.46	1,282	0.45	1,275
RRpassage			0.02	1,1459	0.02	1,1448
RRperson			7.08*	1,306	1.47	1,293
Rspan					7.70*	1,276

* $p < .05$

Table 11. Random variance components for mixed models. Percentage Important Details.

	Model 1		Model 2		Model 3	
	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	496.80*	17.20	496.80*	17.20	496.80*	17.20
Residual	253.27*	9.61	238.40*	9.04	234.21*	8.93
Intercept	153.48*	16.66	152.77*	16.38	140.80*	15.53
Overall R^2	.49	-	.52	-	.69	-
Fixed-effect R^2	.18	-	.21	-	.41	-

Var = variance; * $p < .05$

Table 12. *F* tests for fixed effects of passage, age, reading goal, rereading, and working memory on overall summary quality.

	Model 1		Model 2		Model 3	
	F	*df*	*F*	*df*	*F*	*df*
Passage	75.90*	5,1385	75.66*	5,1391	76.80*	5,1379
Age	14.31*	1,279	12.54*	1,285	14.18*	1,280
Passage x Age	3.31*	5,1385	2.86*	5,1390	3.24*	5,1378
Reading Goal	0.26	1,279	0.36	1,284	0.56	1,279
Age X RG	0.00	1,279	0.00	1,282	0.07	1,276
RRpassage			0.00	1,1453	0.26	1,1444
RRperson			1.80	1,306	0.00	1,294
Rspan					15.25*	1,276

* $p < .05$

Table 13. Random variance components for mixed models. Overall Summary Quality.

	Model 1		Model 2		Model 3	
	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	1.37*	.05	1.37*	.05	1.37*	.05
Residual	.72*	.03	.69*	.03	.68*	.03
Intercept	.44*	.05	.44*	.05	.39*	.04
Overall R^2	.47	-	.50	-	.50	-
Fixed-effect R^2	.15	-	.18	-	.22	-

Var = variance; * $p < .05$

Table 14. *F* tests for fixed effects of passage, age, reading goal, rereading, and working memory on Multiple Choice Scores.

	Model 1		Model 2		Model 3	
	F	*df*	*F*	*df*	*F*	*df*
Passage	26.33*	5,1395	26.03*	5,1401	26.87*	5,1389
Age	29.98*	1,279	28.33*	1,285	27.26*	1,280
Passage x Age	6.36*	5,1395	7.04*	5,1400	6.79*	5,1388
Reading Goal	1.84	1,279	2.11	1,284	2.22	1,280
Age X RG	0.15	1,279	0.10	1,281	0.01	1,276
RRpassage			0.14	1,1463	0.38	1,1457
RRpassage x RG			5.30*	1,1453	5.66*	1,1451
RRpassage x RG x Age					4.36*	1,1422
RRperson			0.59	1,305	0.24	1,294
Rspan					27.20*	1,277

* $p < .05$

Table 15. Random variance components for mixed models. Multiple Choice Scores.

	Model 1		Model 2		Model 3	
	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	3.09*	.11	3.09*	.11	3.09*	.11
Residual	1.69*	.06	1.64*	.06	1.62*	.06
Intercept	1.06*	.11	1.07*	.11	.90*	.10
Overall R^2	.45	-	.47	-	.48	-
Fixed-effect R^2	.11	-	.12	-	.18	-

Var = variance; * $p < .05$

Judgment Magnitude

Both prediction and postdiction confidence varied as a function of age for main ideas, important details, and multiple choice (not for themes or overall summaries)—younger adults had greater confidence than older adults (See Table 16). There were also significant individual differences in judgment magnitude (See Table 17) but neither prediction nor postdiction confidence varied as a function of reading goal for any measure.

Rereading and working memory capacity had inconsistent effects on judgment magnitude. As rereading increased among passages, predictions about overall summary quality [$F(1,1433) = 7.37$, $p < .01$] and multiple choice scores [$F(1,1424) = 9.02$, $p < .01$] decreased. There was no effect of person-level rereading on prediction magnitude. As working memory increased, average predictions about main ideas [$F(1,279) = 6.71$, $p = .01$] and multiple choice scores [$F(1,278) = 4.06$, $p = .04$] increased. The postdiction pattern was somewhat consistent with the prediction pattern. Again, as working memory increased, average predictions about main ideas [$F(1,277) = 6.35$, $p = .01$] and multiple choice scores [$F(1,281) = 7.76$, $p = .01$] increased. However, there was no longer any effect of passage-level rereading on judgments, but as person-level rereading increased, postdictions for overall summary quality decreased [$F(1,290) = 4.30$, $p = .04$].

Table 16. Means for prediction and postdiction magnitude by age for each comprehension measure by age.

	Young	Old	
Pre-Theme	5.11(.08)	5.04 (.08)	$F(1,282) = .22$
Post-Theme	4.73 (.08)	4.95 (.09)	$F(1,282) = 3.93$*
Pre-MI	66.18 (1.59)	54.91 (1.62)	$F(1,280) = 23.02$**
Post-MI	59.48 (1.64)	52.04 (1.66)	$F(1,278) = 9.58$*
Pre-ID	60.07 (1.67)	53.10 (1.69)	$F(1,279) = 7.93$*
Post-ID	54.64 (1.71)	49.89 (1.73)	$F(1,277) = 3.47$
Pre-Overall	4.93 (.08)	4.87 (.08)	$F(1,281) = .16$
Post-Overall	4.40 (.09)	4.48 (.09)	$F(1,281) = .53$
Pre-MC	6.06 (.09)	5.71 (.09)	$F(1,279) = 6.34$*
Post-MC	4.99 (.08)	4.74 (.08)	$F(1,281) = 3.63$

Note: Means with standard error in parentheses. * $p < .05$; ** $p < .001$

Table 17. Random variance of the intercept (individual differences) for prediction and postdiction magnitude.

	Variance	*SE*
Pre-Theme	.72	.07**
Post-Theme	.83	.08**
Pre-MI	322.51	29.71**
Post-MI	344.89	31.65**
Pre-ID	360.55	32.75**
Post-ID	378.09	34.26**
Pre-Overall	.72	.07**
Post-Overall	.88	.09**
Pre-MC	.99	.09**
Post-MC	.81	.08**

Note: ** $p < .001$.

Metacomprehension

Absolute Accuracy (Bias)

See Table 18 for bias scores by age and reading goal. See Table 19 for aggregate bias scores and change in aggregate bias scores from prediction to postdiction.

Themes: Participants were underconfident on average predicting their theme quality [$t(1667) = -2.14$, $p = .03$, $d = -0.10$]. There was no main effect of age or reading goal on pre-theme bias, but they interacted significantly [$F(1,1389) = 7.19$, $p < .01$, $d = 0.65$]. Younger adults were more underconfident than older adults for professors but older adults were more underconfident than younger adults for acquaintances. Participants were also underconfident on average postdicting their theme quality [$t(1667) = -7.61$, $p < .0001$, $d = - 0.37$]. There was a main effect of age on post-theme bias [$F(1,1389) = 6.26$, $p = .01$, $d = - 0.31$], with younger adults showing more underconfidence than older adults. Thus, younger adults tended to be more underconfident overall, particularly for post-themes, but older adults predicting themes in the acquaintance condition showed the most underconfidence.

Main Ideas (MI): Participants were on average overconfident predicting the percentage of main ideas they would include in their summaries [$t(1668) = 9.37$, $p < .0001$, $d = 0.46$]. There was no effect of reading goal on main idea prediction bias, but there was a significant effect of age, [$F(1,1390) = 4.93$, $p = .03$, $d = 0.27$], with younger adults showing more overconfidence than older adults. Participants were also on average overconfident postdicting their main ideas [$t(1660) = 4.00$, $p < .0001$, $d = 0.20$]. There was a significant age by reading goal interaction for main idea postdiction bias [$F(1,1382) = 5.70$, $p = .02$, $d = 0.57$], such that younger adults were more overconfident

with acquaintances but older adults were more overconfident with bosses. Thus, younger adults were more overconfident than older adults initially, but reading goal moderated this effect at test.

Important Details (ID): Participants were, on average, overconfident predicting the percentage of important details [$t(1668) = 39.12$, $p < .0001$, $d = 1.92$] they would include in their summaries. There was no main effect of age or reading goal on important detail prediction bias, but there was a significant age by reading goal interaction, [$F(1390) = 5.07$, $p = .02$, $d = 0.54$]. Whereas younger adults were more overconfident with acquaintances, older adults were more overconfident with bosses. This effect maintained for important detail postdiction bias [$F(1,1371) = 6.60$, $p = .01$, $d = 0.62$], with similar differences in magnitude. As with predictions, the sample was also overconfident postdicting important details [$t(1649) = 34.74$, $p < .0001$, $d = 1.71$]. In summary, reading goal robustly moderated age differences in overconfidence about important details, obtaining before and after test.

Overall Summary: Participants were overconfident, on average, predicting their overall summary quality [$t(1662) = 28.43$, $p < .0001$, $d = 1.39$]. There was no effect of reading goal on overall summary prediction bias, but there was a significant age difference [$F(1, 1384) = 4.32$, $p = .04$, $d = -0.25$], with older adults showing more overconfidence than younger adults. This effect maintained for post-overall bias [$F(1,1370) = 8.12$, $p = .004$, $d = -0.33$]. As with predictions, participants were also overconfident, on average, postdicting their summary quality [$t(1648) = 15.50$, $p < .0001$, $d = 0.76$]. Thus, older adults were robustly overconfident in their overall summarizing ability than younger adults.

Multiple Choice (MC): Participants were on average overconfident predicting their multiple choice scores [$t(1673) = 29.42$, $p < .0001$, $d = 1.44$] There was no effect of reading goal on multiple choice predictions, but there was a significant age difference [$F(1,1395) = 6.85$, $p = .01$, $d = -0.32$], with older adults showing more overconfidence than younger adults. This effect maintained for multiple choice postdiction bias [$F(1,1381) = 13.64$, $p < .001$, $d = -0.45$]. As with predictions, participants were also overconfident overall postdicting multiple choice scores [$t(1659) = 8.46$, $p < .0001$, $d = 0.42$]. Thus, older adults were robustly overconfident in their ability to answer multiple choice questions than younger adults.

Neither rereading nor working memory predicted average bias scores, with two exceptions. Rereading predicted overall summary pre-Bias (as rereading increased, bias increased) and working memory predicted multiple choice pre-Bias (as working memory increased, bias decreased). Average bias scores changed significantly from predictions to postdictions for all dependent measures. Participants became less calibrated on themes, but better calibrated on all other measures.

To summarize, participants were not well calibrated, showing significantly non-zero bias scores in the aggregate and as a function of age group and reading goal—they tended to be slightly underconfident with themes but overconfident on all other measures. The most robust and consistent findings in bias—occurring for predictions and postdictions, an effect not moderated by reading goal—were that broad measures of comprehension (overall summary quality and multiple choice) revealed older adults to be more overconfident than younger adults. Another pattern, as seen with main ideas and important details, is that overconfidence was a larger issue for younger adults when

summarizing for acquaintances but a larger issue for older adults when summarizing for bosses. There was no effect of reading goal on postdiction bias for any measure. Postdictions were significantly less biased than predictions except for themes, where the effect was reversed.

Table 18. Bias scores by age group and reading goal.

	Young		Old	
	Professor	Acquaintance	Boss	Acquaintance
Pre-Theme	-0.25 (.14)	-0.03 (.14)	0.21 (.14)	-0.32 (.14)*
Post-Theme	-0.59 (.15)*	-0.46 (.14)*	0.01 (.15)	-0.33 (.15)*
Pre-MI	9.31 (2.56)*	12.08 (2.43)*	7.87 (2.52)*	2.38 (2.52)
Post-MI	2.32 (2.56)	6.58 (2.44)*	6.00 (2.53)*	-1.74 (2.53)
Pre-ID	27.41 (2.89)*	31.34 (2.75)*	34.41 (2.85)*	25.59 (2.84)*
Post-ID	22.82 (2.84)*	25.82 (2.70)*	31.86 (2.81)*	20.52 (2.81)*
Pre-Overall	0.87 (.13)*	1.00 (.12)*	1.31 (.13)*	1.08 (.13)*
Post-Overall	0.32 (.14)*	0.50 (.13)*	0.94 (.14)*	0.66 (.14)*
Pre-MC	1.12 (.16)*	1.27 (.15)*	1.78 (.15)*	1.42 (.15)*
Post-MC	0.07 (.14)	0.15 (.14)	0.83 (.14)*	0.44 (.14)*

Note: Means with standard errors in parentheses. * = significantly different from zero ($p < .05$)

Table 19. Aggregate bias scores and change in aggregate bias scores from prediction to postdiction.

	Overall	Pre-Post Change
Pre-Theme	-0.09 (.04)	
		$t(1667) = 9.63$, $p < .0001$, $d = 0.47$
Post-Theme	-0.34 (.04)	
Pre-MI	7.98 (.86)	
		$t\ (1660) = 9.12$, $p < .0001$, $d = 0.45$
Post-MI	3.39 (.85)	
Pre-ID	29.75 (.76)	
		$t\ (1649) = 8.60$, $p < .0001$, $d = 0.42$
Post-ID	25.40 (.73)	
Pre-Overall	1.07 (.04)	
		$t\ (1648) = 14.70$, $p < .0001$, $d = 0.72$
Post-Overall	0.60 (.04)	
Pre-MC	1.40 (.05)	
		$t\ (1659) = 28.47$, $p < .0001$, $d = 1.40$
Post-MC	0.37 (.04)	

Note: Means with standard errors in parentheses. * = significantly different from zero ($p < .05$)

Relative Accuracy

Relative metacomprehension accuracy was measured by within-person rank-order correlations (Stuart's tau-c) of metacomprehension variables with performance across the six passages. Participants were able to discriminate among passages more and less well understood, by all measures of comprehension, although correlations were small. For the most part, there were no age differences or differences due to reading goal in relative judgment accuracy (See Table 19). However, there were a few exceptions. Those in the acquaintance condition (tau-c = .10) were better able to discriminate among their themes at study than those in the professor/boss condition (tau-c = -.003), [$F(1,264) = 4.78$, $p =$

.03, d = - 0.26]. There was a significant age by reading goal interaction in the ability to predict relative main idea percentages [$F(1,267) = 6.20$, $p = .01$, $d = 0.60$] and relative overall summary quality, [$F(1,265) = 4.62$, $p = .03$, $d = 0.50$]. In both cases, younger adults were better than older adults when reading for an acquaintance but older adults were better than younger adults when reading for a boss. There was just one age difference in postdictions--when postdicting relative main idea percentages, younger adults were better, [$F(1,267) = 4.77$, $p = .03$, $d = 0.28$]. Neither rereading nor working memory significantly influenced relative judgment accuracy, with the exception of important detail relative prediction accuracy--as rereading increased the relative judgment accuracy of important details decreased [$F(1,252) = 4.29$, $p = .04$]. Relative postdictions were significantly more accurate than relative predictions for important details, overall summary, and multiple choice scores, but not for themes or main ideas (See Table 20).

Table 19. Relative judgment accuracy (t_c) by age and reading condition.

	Young		Old	
	Professor	Acquaintance	Boss	Acquaintance
Pre-Theme	-.02 (.05)	.06 (.05)	.01 (.05)	.15 (.05)*
Post-Theme	.06 (.05)	.09 (.05)	.03 (.05)	.15 (.05)*
Pre-MI	-.01 (.05)	.10 (.04)*	.14 (.05)*	.02 (.05)
Post-MI	.09 (.04)	.13 (.05)*	-.06 (.05)	.07 (.05)
Pre-ID	.05 (.05)	.11 (.04)*	.15 (.05)*	.10 (.05)*
Post-ID	.17 (.04)*	.19 (.04)*	.20 (.04)*	.30 (.04)*
Pre-Overall	.03 (.05)	.17 (.05)*	.18 (.05)*	.11 (.05)*
Post-Overall	.17 (.05)*	.20 (.05)*	.17 (.05)*	.20 (.05)*
Pre-MC	.07 (.05)	.13 (.05)*	-.001 (.05)	.11 (.05)*
Post-MC	.18 (.05)*	.20 (.04)*	.21 (.05)*	.14 (.05)*

Note: Means with standard errors in parentheses. * = significantly different from zero ($p < .05$)

Table 20. Prediction relative accuracy, postdiction relative accuracy, and change in relative accuracy (t_c).

	Prediction Accuracy	Postdiction Accuracy	Change in Accuracy
Themes	.05 (.02)*	.08 (.03)*	t (263) = -1.11, d = - 0.14
Main Ideas	.06 (.02)*	.06 (.02)*	t (264) = -0.10, d = -0.01
Important Details	.10 (.02)**	.21 (.02)**	t (260) = -3.91, d = -0.48**
Overall Summaries	.12 (.02)**	.18 (.03)**	t (260) = -2.10, d = -0.26*
Multiple Choice	.08 (.02)*	.18 (.02)**	t (245) = -3.48, d = -0.44**

Note: Prediction and postdiction columns report means with standard errors in parentheses. d = Cohen's d effect size. * $p < .05$; ** $p < .001$.

Multilevel Modeling (MLM) of Metacomprehension

See Tables (21-30) for MLM significance tests. Fixed effects and random effects are reported in separate tables. Tables of fixed effects for all dependent measures include 4 models. Model 1 includes passage, age, and reading goal as predictors; Model 2 adds passage-level and person-level predictions; Model 3 adds passage-level and person-level rereading; Model 4 adds working memory capacity. All models included subject as a random effect. Tables of random effects include an unconditional model (person intercept and residual variance), residual, intercept (random effect, or individual differences), Overall R^2 (variance accounted for by fixed and random effects), and Fixed-effect R^2.

As seen in Model 2 of the Random Effects tables, including the random intercept variance of predictions or postdictions (the difference between Overall R^2 and Fixed-

effect R^2) more than doubled the total variance accounted for by the experimental factors (passage, age, reading goal) and judgments for all measures of comprehension except the percentage of main ideas included in summaries. Furthermore, including the random intercept variance, Model 2 accounted for about half of the variance in all measures of comprehension except themes and main ideas. Thus, individual differences in average scores for most measures of comprehension were a substantial source of variance in scores.

There were significant effects of passage for all models but, again, for tractability, these are not described here. In short, age and passage had interactive effects on all measures, with varying effects across dependent measures. Descriptions of these effects and an analysis of possible covariates will be saved for future analyses, beyond this dissertation.

Theme Prediction Accuracy: As seen in Tables 21 and 22, person level differences in average theme predictions were associated with theme scores, but passage-level theme predictions were not (See Model 2). Thus, although those who predicted higher theme scores received them, individuals were not able to discriminate among themes of varying quality.

The effect of age is seen in the interactive effect with passage-level theme prediction, qualified by a 3-way interaction person-level theme prediction (See Model 2). There were individual differences in the degree to which passage-level increases in theme prediction were associated with increases in theme score, with stronger effects for older adults than younger adults. Of note, working memory, but not rereading, partially

mediated the interactive effects of passage level theme prediction by person level theme prediction (see Models 3 and 4).

The effect of reading goal is seen in the 3-way interaction of person-level theme prediction, person-level rereading, and reading goal (see Model 3). Although the relationship between individual differences in theme prediction and theme score was amplified by individual differences in rereading, the effect was larger for the professor/boss condition than for the acquaintance condition.

Table 21. *F* tests for fixed effects of passage, age, theme predictions, reading goal, rereading, and working memory on theme scores.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	26.23*	5,1390	25.43*	5,1405	23.85*	5,1407	23.38*	5,1399
Age	2.25	1,279	1.72	1,297	3.32	1,305	3.44	1,302
Passage x Age	5.14*	5,1390	3.19*	5, 1405	2.91*	5, 1411	2.99*	5,1402
Reading Goal	1.27	1,279	1.01	1,291	1.43	1,298	1.38	1,298
Age x RG	3.34	1,279	4.60*	1, 289	4.85*	1,288	5.70*	1,287
TPpassage			1.77	1,1413	1.84	1,1429	1.87	1,1417
TPperson			6.47*	1,290	9.67*	1,292	7.21*	1,291
TPpassage x Age			4.37*	1,1411	3.90*	1,1409	3.18	1,1397
TPpassage x TPperson			3.30	1,1399	3.99*	1,1413	2.53	1,1402
TPpassage x TPperson x Age			9.93*	1,1403	5.95*	1,1412	5.74*	1,1407
RRpassage					5.71*	1,1503	3.96*	1,1501
RRperson					0.04	1,339	0.17	1,338
TPperson x RRperson x Reading Goal					8.87	1,317	6.58*	1,304
Rspan							6.12*	1,296

* $p < .05$

Table 22. Random variance components for mixed models. Theme score predictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	1.97*	.07	1.97*	.07	1.97*	.07	1.97*	.07
Residual	1.31*	.05	1.27*	.05	1.20*	.05	1.16*	.04
Intercept	.51*	.06	.48*	.06	.46*	.06	.43*	.05
Overall R^2	.34	-	.36	-	.39	-	.41	-
Fixed-effect R^2	.08	-	.11	-	.16	-	.19	-

Var = variance; * $p < .05$

Main Ideas (MI) Prediction Accuracy: As seen in Tables 23 and 24, individual differences in main idea predictions were associated with the percentage of main ideas included in summaries but passage-level effects were only significant controlling on working memory capacity (see Model 4). Age did not moderate these effects in any way.

There were no main effects of reading goal on main idea prediction accuracy but reading goal moderated an interaction between passage-level prediction and passage-level rereading. The relationship of passage-level predictions to percentage of main ideas generated was smaller as rereading across passages increased and smallest for the acquaintance condition. Reading goal also moderated an interaction between person-level main idea prediction and passage-level rereading such that the relationship between person level main idea prediction and percentage of main ideas generated was most amplified for those in the professor/boss condition as they reread more across passages (See Model 3), but not when controlling on working memory capacity (See Model 4). Working memory interacted with passage-level prediction such that the ability to judge main ideas among passages increased with larger working memory.

Table 23. *F* tests for fixed effects of passage, age, main ideas prediction, reading goal, rereading, and working memory on percentage of main ideas included in summaries.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	133.96*	5,1391	122.61*	5,1409	115.80*	5,1409	111.98*	5,1399
Age	11.17*	1,279	10.51*	1,299	11.20*	1,301	11.71*	1,296
Passage x age	5.55*	5,1391	5.06*	5,1408	3.95*	5,1410	3.25*	5,1400
RG	0.18	1,279	0.36	1,293	0.62	1,296	0.59	1,300
Age x RG	0.14	1,279	0.05	1,288	0.07	1,285	0.39	1,282
MIpassage			2.95	1,1413	3.39	1,1453	5.19*	1,1462
MIPperson			18.16*	1,290	21.01*	1,292	16.79*	1,290
RRpassage					0.82	1,1536	0.07	1,1522
RRperson					0.00	1,325	0.04	1,316
MIpassage x RRpassage x RG					4.84*	1,1623	4.21*	1,1607
MIPperson x RRpassage x RG					4.06*	1,1480	0.89	1,309
Rspan							12.31*	1,283
MIPassage x Rspan							4.46*	1,1442

* $p < .05$

Table 24. Random variance components for mixed models. Main idea predictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	854.01*	29.56	854.01*	29.56	854.01*	29.56	854.01*	29.56
Residual	508.81*	19.30	493.45*	18.72	460.34*	17.55	445.27*	17.08
Intercept	119.09*	17.59	104.79*	16.16	104.05*	16.07	88.57*	14.74
Overall R^2	.40	-	.42	-	.46	-	.48	-
Fixed-effect R^2	.26	-	.30	-	.34	-	.37	-

Var = variance; * $p < .05$

Important Details (ID) Prediction Accuracy: As seen in Tables 25 and 26, neither passage-level nor person-level predictions were related to the percentage of important details (ID) included in passages, except for passage-level predictions when controlling on rereading. There were age differences in the relationship of passage-level predictions to percentage IDs (stronger effects for younger adults) but which were not significant after controlling on rereading. Controlling on rereading, however, elicited a person-level prediction, person-level rereading, and age interaction such that the strongest relationship between ID predictions and scores is for younger adults who reread more. This effect, however, goes away when controlling on working memory capacity.

Passage-level predictions and person-level rereading interacted with reading goal such that the relationship of passage-level increase to percentage of IDs included in summaries is greatest for those who reread more in the professor/boss condition.

Table 25. *F* tests for fixed effects of passage, age, reading goal, important details prediction, rereading, and working memory on percentage of important details included in summaries.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	93.49*	5,1390	83.17*	5,1404	81.82*	5,1406	78.59*	5,1396
Age	19.45*	1,279	16.37*	1,293	14.05*	1,296	13.82*	1,297
Passage x age	4.70*	5,1390	6.40*	5,1402	7.72*	5,1405	6.57*	5,1396
Reading Goal	0.65	1,279	0.38	1,290	0.30	1,291	0.04	1,293
Age x RG	0.28	1,279	0.73	1,284	1.23	1,281	1.06	1,278
IDPpassage			3.69	1,1401	4.12*	1,1428	1.16	1,1427
IDPperson			2.51	1,285	3.84	1,289	1.52	1,288
IDPpassage x Age			4.49*	1,1404	3.16	1,1407	3.84	1,1410
RRpassage					0.00	1,1498	0.08	1,1484
RRperson					1.67	1,303	0.34	1,299
IDPpassage x RRperson x RG					4.21*	1,1439	5.45*	1,1426
IDPperson x RRperson x Age					4.48*	1,299	2.28	1,289
Rspan							7.17*	1,287
RRpassage x RRperson x Rspan							5.30*	1,1425

* $p < .05$

Table 26. Random variance components for mixed models. Important details predictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	496.80*	17.20	496.80*	17.20	496.80*	17.20	496.80*	17.20
Residual	253.27*	9.61	244.34*	9.27	227.11*	8.64	220.93*	8.44
Intercept	153.48*	16.66	149.46*	16.19	145.07*	15.79	135.05*	14.98
Overall R^2	.49	-	.51	-	.54	-	.56	-
Fixed-effect R^2	.18	-	.21	-	.25	-	.28	-

Var = variance; * $p < .05$

Overall Summary (OS) Prediction Accuracy: As seen in Tables 27 and 28, Passage-level and person-level predictions were related to overall summary quality, showing that individuals predicted which summaries would be of higher overall quality and that individual differences in average predictions were related to individual differences in overall summary quality scores. There were no age differences or differences due to reading goal in the ability to predict the quality of summaries.

Individual differences in rereading moderated both the relationship between passage-level predictions and overall summary quality scores and the relationship between person-level predictions and scores. There was a slight decrease in scores as predictions increased across passages, which became amplified (more negative) as average rereading between individuals increased. Reading goal, however, moderated the interaction of person-level predictions and person-level rereading such that although those with higher predictions tended to score higher, the effect was reduced for those who reread more and reduced further for those in the acquaintance condition. Adding working

memory to the model elicited new 3-way interactions. The 2-way interaction of person-level predictions and person-level rereading is now moderated by passage-level rereading such that the effects are amplified as predictions increase across passages. The 2-way interaction of passage-level predictions and person-level rereading is also affected, by working memory capacity, such that the effect is smaller (less negative) for those with higher working memory capacity.

To summarize, participants could predict the quality of their summaries with age equivalence in this ability. Rereading seemed to be a behavior of those who had more trouble judging the quality of their summaries but less so for those with higher working memory capacity.

Table 27. *F* tests for fixed effects of passage, age, overall summary quality prediction, reading goal, rereading, and working memory on overall summary quality scores.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	75.90*	5,1385	56.09*	5,1399	53.50*	5,1404	54.27*	5,1395
Age	14.31*	1,279	13.49*	1,298	15.65*	1,305	14.63*	1,303
Passage x age	3.31*	5,1385	2.54*	5,1399	2.96*	5,1405	3.55*	5,1396
Reading Goal	0.26	1,279	0.11	1,294	0.31	1,300	0.39	1,303
Age x RG	0.00	1,279	0.12	1,282	0.51	.47	0.79	1,283
OSPpassage			8.90*	1,1393	6.93*	1,1400	7.75*	1,1394
OSPperson			11.49*	1,292	16.37*	1,296	10.75*	1,293
RRpassage					0.09	1,1486	0.30	1,1488
RRperson					1.50	1,321	0.00	1,310
OSPpassage x RRperson					5.24*	1,1413	1.74	1,1403
OSPperson x RRperson					4.71*	1,309	2.17	1,296
OSPperson x RRperson x RG					12.27*	1,309	7.62*	1,294
RRperson x RG							5.01	1,305
OSPperson x RRpassage x RRperson							4.18*	1,1445
Rspan							17.29*	1,306
OSPpassage x RRperson x Rspan							5.57*	1,1383

* $p < .05$

Table 28. Random variance components for mixed models. Overall summary predictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	1.37*	.05	1.37*	.05	1.37*	.05	1.37*	.05
Residual	.72*	.03	.70*	.03	.65*	.02	.63*	.02
Intercept	.44*	.05	.41*	.05	.39*	.04	.34*	.04
Overall R^2	.47	-	.49	-	.53	-	.54	-
Fixed-effect R^2	.15	-	.19	-	.24	-	.29	-

Var = variance; * $p < .05$

Multiple Choice (MC) Prediction Accuracy: As seen in Tables 29 and 30, person-level, but not passage-level, multiple choice predictions were related to individual differences in multiple choice scores. There was a slight increase in the relationship between passage-level predictions and multiple choice scores for those with higher working memory capacity, which was significantly reduced for younger adults. There was no effect of reading goal on ability predict multiple choice scores.

Table 29. *F* tests for fixed effects of passage, age, multiple choice predictions, reading goal, rereading, and working memory on multiple choice scores.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	26.33*	5,1395	23.66*	5,1413	21.75*	5,1417	21.16*	5,1409
Age	29.98*	1,279	35.26*	1,301	37.61*	1,301	35.36*	1,303
Passage x Age	6.36*	5,1395	5.49*	5,1411	5.49*	5,1413	5.41*	5,1406
Reading Goal	1.84	1,279	0.28	1,297	0.34	1,299	0.40	1,304
Age x RG	0.15*	1,279	0.08	1,285	0.37	1,281	0.61	1,282
MCPpassage			2.70	1,1407	3.48	1,1434	2.97	1,1429
MCPperson			32.56*	1,288	38.78*	1,288	29.79*	1,291
RRpassage					0.16	1,1505	0.47	1,1504
RRperson					0.54	1,322	0.01	1,315
RRpassage x RG							5.87*	1,1487
Rspan							17.39*	1,298
MCPpassage x Age x Rspan							5.50*	1,1396

* $p < .05$

Table 30. Random variance components for mixed models. Multiple choice predictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	3.09*	.11	3.09*	.11	3.09*	.11	3.09*	.11
Residual	1.69*	.06	1.63*	.06	1.54*	.06	1.49*	.06
Intercept	1.06*	.11	.89*	.10	.88*	.10	.77*	.09
Overall R^2	.45	-	.47	-	.50	-	.52	-
Fixed-effect R^2	.11	-	.18	-	.22	-	.27	-

Var = variance; * $p < .05$

For Postdictions, see Tables 31-40.

Theme Postdiction Accuracy: As seen in Tables 31 and 32, at the passage level, higher theme postdictions were associated with higher scores. But, although individual differences in theme postdictions were related to theme scores, as postdictions increased, theme scores decreased.

There were no age differences in the ability to postdict the quality of themes, but reading goal had some influence on performance. There was a significant interaction of person-level theme postdictions, person level rereading, and reading goal. There was a slight increase in the relationship of person level postdictions to scores for those who reread more on average, which was significantly amplified for those in the professor/boss condition. Controlling on working memory capacity, there was also a significant interaction of reading goal with passage-level theme postdiction such that the increase in theme scores associated with an increase in postdictions across passage was reduced for those in the professor/boss condition. However, these predictors interacted with working memory capacity. Within an individual, although the increase in the relationship of

passage-level postdiction to score was smaller for those in the professor/boss condition than for those in the acquaintance condition, the effect was less reduced for those with higher working memory.

In summary, participants showed some ability to postdict their theme quality, with age equivalence. Rereading and higher working memory capacity had some benefit to the judgment ability of those in the professor/boss condition.

Table 31. *F* tests for fixed effects of passage, age, reading goal, theme postdictions, rereading, and working memory on theme scores.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	26.23*	51,1390	25.48*	5,1405	22.97*	5,1409	21.78*	5,1399
Age	2.25	1,279	1.40	1,296	1.87	1,298	2.17	1,297
Passage x age	5.14*	5,1390	3.64*	5,1405	3.52*	5,1407	3.16*	5,1397
Reading Goal	1.27	1,279	1.58	1,294	2.06	1,294	2.00	1,299
Age x RG	3.34	1,279	4.18*	1,286	4.46*	1,282	5.47*	1,281
TPpassage			7.64*	1,1399	9.06*	1,1420	7.08*	1,1404
TPperson			7.64*	1,295	8.99*	1,292	6.85*	1,290
RRpassage					3.87*	1,1514	4.01*	1,1502
RRperson					0.00	1,329	0.41	1,313
TPperson x RRperson x RG					5.45*	1,311	4.33*	1,298
TPpassage x RG							3.99*	1,1408
Rspan							5.92*	1,297
TPpassage x Rspan x RG							4.67*	1,1391
RRpassage x Rspan							7.23*	1,1443

* $p < .05$

Table 32. Random variance components for mixed models. Theme postdictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	1.97*	.06	1.97*	.06	1.97*	.06	1.97*	.06
Residual	1.31*	.05	1.26*	.05	1.20*	.05	1.15*	.04
Intercept	.51*	.06	.48*	.06	.47*	.06	.44*	.06
Overall R^2	.34	-	.36	-	.39	-	.42	-
Fixed-effect R^2	.10	-	.12	-	.15	-	.19	-

Var = variance; * $p < .05$

Main Idea (MI) Postdiction Accuracy: As seen in Tables 33 and 34, passage-level and person-level main idea (MI) postdictions predicted the percentage of main ideas included in summaries. Within and between individuals, as postdictions increased, so did MI scores. There were no age differences in the ability to postdict MI scores but reading goal, rereading, and working memory capacity had some influence.

Reading goal interacted with passage-level and person-level MI postdiction such that although scores increased with increases in postdictions across passages, the effect was smaller for those who gave higher postdictions on average, and smallest for those in the acquaintance condition. Reading goal also interacted with person-level postdiction and person-level rereading. The tendency for those who reread more to show a stronger relationship of main idea postdictions to percentage of main ideas included in summaries was amplified in those in the professor/boss condition. These effects, however, were mediated by working memory capacity.

With working memory in the model, passage-level rereading interacted with passage-level postdictions and person-level postdictions. The effect of within-person

increases in postdictions on percentage of main ideas included in summaries was smallest for those who gave higher postdictions on average as they reread more across passages. Additionally, working memory interacted with person level postdictions and passage level rereading. The relationship of individual differences in main idea postdictions to main idea scores increases with more rereading across passages, but the effect is reduced as working memory capacity increases.

In summary, participants could judge the percentage of main ideas included in summaries, with age equivalence in postdiction acuracy. Although rereading is of some benefit to the success of main idea postdiction, particularly when reading for a professor/boss, these effects seem to be largely carried by working memory capacity.

Table 33. *F* tests for fixed effects of passage, age, main ideas postdictions, reading goal, rereading, and working memory on percentage of main ideas included in summaries.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	133.96*	5,1391	130.06*	5,1400	124.48*	5,1408	122.77*	5,1398
Age	11.17*	1,279	11.31*	1,296	11.45*	1,306	11.30*	1,302
Passage x age	5.55*	5,1391	5.63*	5,1400	4.07*	5,1408	4.16*	5,1403
Reading Goal	0.18	1,279	0.40	1,295	0.54	1,298	0.65	1,302
Age x RG	0.14	1,279	0.16	1,282	0.22	1,279	0.43	1,278
MIPpassage			29.48*	1,1407	28.61*	1,1428	21.21*	1,1430
MIPperson			20.49*	1,289	23.81*	1,286	19.10*	1,286
MIPpassage x MIPperson x RG			4.22*	1,1403	4.93*	1,1401	3.84	1,1405
RRpassage					0.23	1,1541	0.04	1,1545
RRperson					0.02	1,395	0.32	1,361
MIPpassage x MIPperson x RRpassage					4.20	1,1613	5.18*	1,1619
MIPperson x RRperson x RG					5.26*	1,304	2.46	1,308
Rspan							8.56*	1,300
MIPperson x RRpassage x Rspan							8.38*	1,1423

* $p < .05$

Table 34. Random variance components for mixed models. Main idea postdictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	854.01*	29.56	854.01*	29.56	854.01*	29.56	854.01*	29.56
Residual	508.81*	19.30	482.63*	18.36	454.81*	17.38	438.81*	16.84
Intercept	119.09*	17.59	103.21*	15.91	97.76*	15.45	81.38*	14.00
Overall R^2	.40	-	.43	-	.47	-	.49	-
Fixed-effect R^2	.26	-	.31	-	.35	-	.39	-

Var = variance; * $p < .05$

Important Detail (ID) Postdiction Accuracy: As seen in Tables 35 and 36, passage-level and person-level important details postdictions predict the percentage of important details included in summaries. There were age differences in both of these abilities, with stronger effects for younger adults than older adults. However, working memory interacted with passage-level postdictions and age such that younger adults with higher working memory capacity showed the largest increase in association between passage level postdictions and percentage of important details included in summaries.

There was also a passage level postdiction by reading goal interaction such that the associated increases in postdictions and scores were larger for those in the acquaintance condition than for those in the professor/boss condition.

There was an interaction of person level ID postdictions and person level rereading such that decrease in ID scores as average ID postdictions increased was reduced for those who reread more. But this was qualified by a 3-way interaction with passage level posdictions, such that the effect was reduced with passage-level increases

in postdictions. These interactions, however, were no longer significant when controlling on working memory capacity, indicating that working memory mediated the effects.

In summary, participants could judge the percentage of important details included in their summaries, but younger adults, particularly those with higher working memory capacity, had superior ability to do so. Rereading and reading for an acquaintance had some benefit.

Table 35. *F* tests for fixed effects of passage, age, reading goal, important details postdictions, rereading, and working memory on percentage of important detail included in summaries.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	93.49*	5,1390	73.90*	5,1386	76.71*	5,1393	75.38*	5,1384
Age	19.45*	1,279	18.65*	1,296	21.26*	1,303	20.07*	1,299
Passage x age	4.70*	5,1390	5.33*	5,1387	6.60*	5,1392	6.10*	5,1386
Reading Goal	0.65	1,279	0.06	1,292	0.01	1,297	0.05	1,294
Age x RG	0.28	1,279	0.58	1,281	1.87	1,282	1.97	1,279
IDPpassage			41.92*	1,1390	26.73*	1,1393	25.71*	1,1392
IDPperson			9.34*	1,291	13.38*	1,294	9.12*	1,293
IDPpassage x Age			5.21*	1,1384	6.45*	1,1395	6.05*	1,1388
IDPpassage x RG			8.28*	1,1381	7.25*	1,1381	7.00*	1,1376
IDPperson x Age			10.36*	1,281	11.09*	1,281	14.25*	1,279
RRpassage					0.09	1,1474	0.11	1,1466
RRperson					11.45*	1,365	4.95*	1,336
IDPperson x RRperson					4.06*	1,313	3.85	1,314
IDPpassage x IDPperson x RRperson					4.35*	1,1426	2.54	1,1402
Rspan							5.05*	1,294
IDPpassage x Rspan x Age							5.16*	1,1369

* $p < .05$

Table 36. Random variance components for mixed models predicting. Important detail postdictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	496.80*	17.20	496.80*	17.20	496.80*	17.20	496.80*	17.20
Residual	253.27*	9.61	236.39*	9.03	219.84*	8.42	212.10*	8.15
Intercept	153.48*	16.66	144.76*	15.73	138.07*	15.11	128.79*	14.32
Overall R^2	.49	-	.52	-	.56	-	.57	-
Fixed-effect R^2	.18	-	.23	-	.28	-	.32	-

Var = variance; * $p < .05$

Overall Summary (OS) Postdiction Accuracy: As seen in Tables 37 and 38, passage level and person level overall summary quality postdictions predicted overall summary quality scores, but age did not moderate either of these effects.

Rereading and reading goal had some effect on overall summary quality score judgments. Passage level rereading and person level rereading had an interactive effect on passage level overall postdictions. Although the increase in postdictions across passages was associated with an increase in overall scores more as rereading increased across passages, the effect was reduced for individuals who reread less on average, except when controlling on working memory. Person level rereading amplified the effect of passage level increases in overall postdictions on overall summary scores, which was further amplified in the professor/boss condition. The same amplification effects occurred for person level postdiction effects except when controlling on working memory.

Working memory mediated the interaction of passage level postdiction and person level rereading, the interaction of person level postdiction, person level rereading and reading goal, and the interaction of passage-level postdiction, passage level rereading,

and person level rereading. Also, working memory interacted with passage-level rereading. Within an individual, rereading's effects on overall summary score was amplified for those with higher working memory capacity.

In summary, participants could judge the quality of their summaries with age equivalence. Rereading and reading goal had some effect on this ability but their effects were largely mediated by working memory capacity.

Table 37. *F* tests for fixed effects of passage, age, reading goal, overall summary quality postdictions, rereading, and working memory on overall summary quality.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	75.90*	5,1385	60.36*	5,1384	58.50*	5,1391	59.02*	5,1385
Age	14.31*	1,279	13.79*	1,291	13.90*	1,295	14.66*	1,293
Passage x age	3.31*	5,1385	2.87*	5,1384	2.57*	5,1388	2.94*	5,1382
Reading Goal	0.26	1,279	0.46	1,291	0.30	1,294	0.45	1,294
Age x RG	0.00	1,279	0.14	1,280	0.45	1,279	0.87	1,277
OSPpassage			60.20*	1,1380	60.96*	1,1387	48.89*	1,1381
OSPperson			10.55*	1,290	12.05*	1,291	8.95*	1,290
RRpassage					0.55	1,1472	0.26	1,1467
RRperson					4.15*	1,345	0.91	1,335
OSPpassage x RRperson					3.92*	1,1422	2.09	1,1408
OSPpassage x RRperson x RG					4.47*	1,1393	5.71*	1,1394
OSPperson x RRperson x RG					5.51*	1,302	2.92	1,299
OSPpassage x RRpassage x RRperson					4.57*	1,1550	2.13	1,1521
Rspan							10.83*	1,291
RRpassage x Rspan							4.40*	1,1410

* $p < .05$

Table 38. Random variance components for mixed models. Overall summary quality postdictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	1.37*	.05	1.37*	.05	1.37*	.05	1.37*	.05
Residual	.72*	.03	.67*	.03	.62*	.02	.60*	.02
Intercept	.44*	.05	.42*	.05	.40*	.04	.35*	.04
Overall R^2	.47	-	.51	-	.55	-	.56	-
Fixed-effect R^2	.15	-	.20	-	.26	-	.31	-

Var = variance; * $p < .05$

Multiple Choice (MC) Postdiction Accuracy: As seen in Tables 39 and 40, both passage level and person level multiple choice postdictions were related to multiple choice scores. Reading goal did not affect the ability to judge multiple choice scores but age had some interactive effects.

There was a significant interaction of passage-level postdiction, passage-level rereading, and age. As postdictions increased across passages so did multiple choice scores but less so as more rereading occurs and least for older adults. With working memory in the model, age moderates the interactive effect of passage-level and person-level postdictions on multiple choice scores. The positive passage-level relationship between postdictions and multiple choice scores was strongest for younger adults who gave higher average postdictions.

In summary, participants could judge the number of multiple choice questions they answered correctly. Older adults who reread more seemed to have the lowest ability to make these judgments and younger adults who tended to give higher judgments seemed to have the highest ability to do so.

Table 39. *F* tests for fixed effects of passage, age, reading goal, multiple choice postdictions, rereading, and working memory on multiple choice scores.

	Model 1		Model 2		Model 3		Model 4	
	F	*df*	*F*	*df*	*F*	*df*	*F*	*df*
Passage	26.33*	5,1395	23.16*	5,1392	22.09*	5,1397	22.74*	5,1392
Age	29.98*	1,279	34.14*	1,289	34.58*	1,291	36.29*	1,292
Passage x age	6.36*	5,1395	3.71*	5,1392	4.13*	5,1396	4.00*	5,1386
Reading Goal	1.84	1,279	0.44	1,287	0.52	1,290	0.47	1,293
Age X RG	0.15	1,279	0.61	1,284	0.89	1,282	1.54	1,280
MCPpassage			65.32*	1,1388	59.35*	1,1389	57.55*	1,1379
MCPperson			57.76*	1,289	53.58*	1,288	54.32*	1,285
RRpassage					0.23	1,1456	0.44	1,1470
RRperson					0.68	1,326	0.12	1,318
MCPpassage x RRpassage x Age					7.39*	1,1550	7.33*	1,1552
MCPpassage x MCPperson x Age							4.12*	1,1374
RRpassage x Age x RG							5.30*	1,1439
Rspan							15.48*	1,288
RRperson x Rspan x RG							10.28*	1,311

* $p < .05$

Table 40. Random variance components for mixed models. Multiple choice postdictions.

	Model 1		Model 2		Model 3		Model 4	
	Var	*SE*	Var	*SE*	Var	*SE*	Var	*SE*
Unconditional	3.09*	.11	3.09*	.11	3.09*	.11	3.09*	.11
Residual	1.69*	.06	1.57*	.06	1.48*	.06	1.44*	.06
Intercept	1.06*	.11	.82*	.09	.85*	.09	.65*	.08
Overall R^2	.45	-	.49	-	.52	-	.53	-
Fixed-effect R^2	.11	-	.23	-	.25	-	.32	-

Var = variance; * $p < .05$

Summary of Results

Vocabulary, Working Memory, and Survey Measures

Older adults had higher vocabulary scores but lower working memory capacity than younger adults. Older adults also reported more reading and summarizing per week in their daily lives than younger adults, and rated the passages in the study to be more familiar, more interesting, and less difficult. Participants showed reduced reading self-efficacy and perceived metacomprehension abilities after task and their posttest reports were more highly correlated with their measured comprehension. As measured by the MCS, younger adults tended to have lower self-reported metacomprehension than older adults, but there were no age differences in reading self-efficacy. The effect of reading goal was all but non-existent.

Passage Comprehension

There were significant individual differences in comprehension scores as well as significant and sizeable age differences (younger adults performed better) on all measures

except themes. Higher working memory was correlated with better passage comprehension but the influence of reading goal and rereading was relatively minor.

Judgment Magnitude

Participants were moderately confident in their comprehension but showed significant individual differences. There were age differences in both prediction and postdiction judgment magnitude for main ideas, important details, and multiple choice scores, with younger adults showing greater confidence than older adults, but not for themes or overall summary quality. Reading goal had no effect. Rereading and working memory had inconsistent effects on judgment magnitude.

Metacomprehension

Aggregate. As measured by bias and tau-c, respectively, participants were generally overconfident and not especially good (but able) at discriminating among more and less well understood passages. Tau-c analyses generally suggested better relative accuracy than MLM for predictions, although correlations were relatively small. MLM showed that overall participants could judge their comprehension by all measures, but more often revealed good person-level than passage-level metacomprehension accuracy, with the latter improving at postdiction. MLM analysis also revealed that significant individual differences in average predictions and postdictions on almost all measures were associated with individual differences in their respective comprehension scores--those who judged higher scores achieved them. Consistent across all statistical approaches, postdictions were generally more accurate than predictions.

Effects of Age, Reading Goal, Rereading and Working Memory. To some extent, different analyses revealed divergent effects on metacomprehension. Bias scores showed

an inconsistent effect of age on calibration but a robust effect of overconfidence for older adults' judgments of overall summary quality and multiple choice scores. A main effect of age on relative accuracy only occurred for main idea postdiction accuracy, with younger adults demonstrating superiority. MLM showed that age only had an effect on predictive accuracy for themes, important details, and multiple choice scores, and not on main ideas or overall summary quality. Older adults' held an advantage with passage-level themes and passage level multiple choice scores (if had higher working memory), but younger adults' advantage was with passage-level important details, primarily when they took advantage of the ability to reread and/or had higher working memory capacity. For the most part, there was age equivalency on postdictive accuracy. Although bias and tau-c did not indicate that rereading and working memory were important influences on metacomprehension accuracy, MLM analysis suggested otherwise. Rereading appears to be something that younger adults, particularly those with higher working memory capacity, may use to excel, in some cases, but which older adults who are struggling may employ to no avail.

The effect of reading goal also depended on the statistical analysis. Whereas bias indicated that younger adults were more often poorly calibrated (overconfident) for acquaintances than professors and older adults more often poorly calibrated (overconfident) for bosses than acquaintances, tau-c showed that relative accuracy was more often better for younger adults in the acquaintance condition than in the professor condition, but better for older adults in the boss condition than in the acquaintance condition. According to MLM, there was no interaction of age and reading goal on metacomprehension accuracy but reading goal did have some effect on all measures

except multiple choice judgment accuracy. Those in the professor/boss condition tended to show some superiority mainly at the person-level but also at the passage-level. And, often, when those in the professor/boss condition reread more they were better able to judge their comprehension than those in the acquaintance condition, but this advantage was largely working memory dependent.

CHAPTER 4

DISCUSSION

There were several unique features and contributions of this study. It is the first to 1) use summary quality as the criterion for metacomprehension judgments and show that people have some ability to judge the quality of their summaries but with much room for improvement; 2) examine age-differences in metacomprehension as measured by oral summaries and reveal that, by traditional statistical measures, older adults are somewhat more overconfident than younger adults, but have fairly comparable abilities to younger adults in relative accuracy 3) allow selective rereading for both younger and older adults and show that it does not mediate age differences but may allow older adults to manage the task and younger adults to excel; 4) examine the influence of a professional versus a social reading goal on older adults comprehension and metacomprehension and show that these reading goals had fairly mild effects, and influenced younger and older adults in different ways; 5) examine change in reading self-efficacy and self-reported metacomprehension and show that they can both be reduced with a metacomprehension task and increase in correlation to actual comprehension as a result; and 6) analyze metacomprehension data with multi-level modeling and compare the results to those from traditional metacomprehension analyses, showing that somewhat different conclusions can result from the analysis used.

The present study was designed to investigate metacomprehension ability measured with summaries and the influence of age, reading goal, and working memory on this ability. It also compared conclusions about metacomprehension ability drawn from traditional metacognitive analyses to those drawn from MLM. Overall, participants showed some metacomprehension ability, but with much room for improvement, and

significant individual differences. In many instances, older adults did not show age-related metacomprehension deficits, although they were more often overconfident than younger adults. Traditional metacognitive analyses suggested age equivalency in within-person accuracy but MLM suggested an age-related deficit in within-person accuracy. Compared to traditional metacognition statistics, MLM revealed a smaller influence of reading goal but a more important influence of working memory on metacomprehension accuracy. Task experience induced more accurate online judgments and more consistency between general subjective and objective metacomprehension abilities. Thus, improving upon the levels of metacomprehension found in this study appears to be possible.

Age Differences in Comprehension

Prior research suggests that older adults are capable of constructing an adequate situation model (Radvansky & Dijkstra, 2007), yet age differences were found on all measures of passage comprehension in this study except themes. The difference may be attributable to how the situation model was measured. Radvansky and Dijkstra inferred older adults were as competent as younger adults at creating a situation model because they confused sentence pairs that were situationally consistent more than ones that were not with rates equal to younger adults. Previous research on text comprehension as measured with summaries is very limited, but the present results are consistent with the majority of earlier studies finding declines with age (Adams, 1991; Byrd, 1985). Further research is needed, however, because previous research involved a variety of text types and measured comprehension in different ways (Johnson, 2003; Madden & Dijkstra, 2010; Radvansky, 1999; Radvansky & Dijkstra, 2007; Radvansky, Zwaan, Curiel, & Copeland, 2001; Stine-Morrow et al., 2006)

Older adults performed worse than younger adults on most measures of comprehension despite having many characteristics that might have supported their comprehension. The older adults in this study were more educated and had higher vocabulary scores than the younger adults, but previous research has shown that education and vocabulary do not necessarily buffer against declines in text comprehension related to aging (Hartley, 1986, 1993b; Johnson, 2003). Older adults also reported having less difficulty with, greater interest in, and greater familiarity with the passages they read than reported by younger adults. Although the empirical validity of self-reports for variables like actual difficulty can be questioned, taken at face value the ratings suggest that the inferior text comprehension of older adults is not likely a function of these variables. In fact, rated passage familiarity did not predict text comprehension and it did not significantly reduce the relationship of working memory capacity to comprehension. This is consistent with previous findings in the literature indicating that domain knowledge moderates the effect of age on text comprehension (Dijkstra, 2001; Jeong & Kim, 2009), but it may not buffer against the consequences of decline in other aspects of cognition, such as working memory and verbal speed (Hambrick & Engle, 2002; Hultsch et al., 1990). The older adults in my sample also reported more reading and summarizing per week in their daily lives than did younger adults, so their inferior text comprehension is not likely to be accounted for by a decrease in practice. Despite having or reporting qualities that may lead one to predict age equivalency in text comprehension, older adults performed worse.

The results demonstrate that age-deficits in text comprehension likely exist, except at the most macro-level of theme generation, but I cannot rule out the possibility

that the age differences were influenced by age differences in episodic memory. Some might claim that using delayed summarization to measure comprehension confounds comprehension with memory, which becomes the principal source of age differences in performance. As such, one may argue that theme quality is the better measure of comprehension than summaries because they were solicited immediately after reading. However, generating a high quality theme also requires less detail integration and could arguably be achieved by extracting the central gist of a passage without the requirement of understanding all components of the passage. Although testing at a delay inherently confounds comprehension with memory, summarization is arguably still an adequate measure of comprehension. Summarization is more dependent on recall than multiple choice performance, which can rely on recognition, yet age effects on summarization and multiple choice performance are similar. For the same reasons, the validity of using summaries to study age differences in comprehension and metacomprehesnion should not be discounted on the premise that summaries confound comprehension with various aspects of summary production such as organization, coherence, and efficiency, even though these are factors in the quality of a summary (Garner, 1985; Kintsch & Kozminsky, 1977; Taylor, 1984). Older adults may achieve equivalent performance to younger adults on theme generation because gist-encoding is better preserved with age, but generating a quality theme may not be sufficient to achieve goals in all areas of life and thus should not be considered a fully adequate demonstration of comprehension.

If there are truly age differences in text comprehension, then what can account for them? Previous research on younger adults showed that working memory is not related to situational model construction (Radvansky & Copeland, 2004) but in the present study

working memory predicts comprehension and older adults were lower in both. The correlation of comprehension with working memory, as measured with the reading span, has been previously established (Carpenter & Just, 1989), but the extent to which working memory accounts for age differences in reading comprehension has been unclear. Lower working memory capacity has been argued to be a prime culprit for reduced text comprehension in older adults (Stine-Morrow et al., 2006), but there is mixed support for a resource deficit account of age differences in text comprehension (Hartley, 1993b; Hultsch et al., 1998; Van der Linden et al., 1999).

Given the extreme age groups design, I made no effort to test statistical mediation of age differences in text comprehension by working memory. However, indirect evidence suggests that working memory could be a culprit for some, but not all, indicators of text comprehension. Although older adults had lower working capacity overall, and higher spans produced higher quality themes, there were no age differences in theme quality. Given that rereading within an individual increased theme scores and older adults reread more than younger adults, older adults may have been able to achieve equivalent theme comprehension in part through rereading. However, on all other measures of comprehension older adults performed worse than younger adults. Even though older adults reread more, their lower working memory capacity may have prevented them from achieving comprehension scores on par with younger adults.

Although there were age-related influences on text comprehension, reading goal did not moderate this effect (reading goal also had no main effect on text comprehension). Although I hypothesized that passages read for an acquaintance may be less well understood because they may be processed more shallowly, this was

unconfirmed. Moreover, although I hypothesized that age differences would be larger for the professor/boss goal because older adults adjust less to experiment goal instructions (Stine-Morrow et al., 2006) and because I expected that older adults' limited cognitive abilities would not allow them to meet the potentially higher standard held in this condition, this was also not borne out. The failure to find an effect of reading goal cannot be presently explained. A weak reading goal manipulation may be responsible, or the processes engaged in the task may simply not be moderated by the audience types. The only prior studies of reading goal effects on older adults' text comprehension have compared processing to storage goals (Stine-Morrow et al., 2006) and recall to comprehension goals. Studies have looked at study vs. entertain goals in younger adults studies but they have shown mixed results (Dudukovic et al., 2004; Kendeou et al., 2011; Linderholm et al., 2008; Linderholm & Wilde, 2010; Narvaez et al., 1999; Van den Broek et al., 2001). Although younger adults report reading texts for study goals more carefully or repeatedly than for entertain goals (Lorch et al., 1993; Narvaez et al., 1999) they do not process texts for study more deeply (Kendeou et al., 2011), nor do they remember more under a study goal (Linderholm et al., 2008). Furthermore, although younger adults may retell stories more accurately and with more detail when asked to focus on accuracy vs. entertainment, there is no difference between groups in comprehension as measured by multiple choice tests (Dudukovic, Marsh, & Tversky, 2004). Thus, participants' perceptions of the effects of reading goal on actual reading may be stronger than the effects of reading goal on actual text comprehension. However, a more comprehensive and ecological exploration of reading goal effects is needed. More

than reading goals, it seems that characteristics of the passages moderate age effects more than audience (e.g. passage by age interactive effects).

Metacomprehension Measured with Summaries

Overall, participants showed some ability to judge most aspects of summary quality, with individual differences, and demonstrated improvements from predictions to postdictions. The low but above-chance accuracy is consistent with previous studies that measured metacomprehension in other ways (Maki, 1998b; Maki et al., 1990) as are the significant individual differences (Chiang et al., 2010). Moreover, the overall superiority of postdictions over predictions is consistent with results from Pierce and Smith (2001) who used gamma correlations to assess the correspondence between judgments and test questions. The superior accuracy of postdictions is also consistent with the metamemory literature (Hertzog & Hultsch, 2000) and may be a consequence of being able to monitor actual output rather than basing judgments on less diagnostic cues such as reading self-efficacy or passage difficulty.

Rereading is a cue that may account for some of the improved metacomprehension accuracy at test. Hwoever, despite rereading not predicting overall quality or multiple choice scores, increased within-person rereading was associated with decreased prediction magnitudes for overall summary quality and multiple choice scores and increased between-person rereading was associated with decreased postdiction magnitude for overall summary quality. This suggests that rereading may not be a diagnostic cue in general, perhaps because it is compensatory behavior in some instances but a labor in vain in others. A question for future study concerns the extent to which

readers know when their rereading will produce a return on investment and when it will not and how this influences the accuracy of metacomprehension judgments.

Different statistical analyses suggest slightly different characterizations of the strength of participants' metacomprehension abilities. Bias scores generally showed overconfidence, moderate for main ideas but more severe for important details, overall summary quality, and multiple choice scores. Tau-c analyses revealed low but statistically significant relative accuracy, with the best accuracy for postdictions of important details, overall summary, and multiple choice scores. MLM partly mirrored these effects, revealing both between and within-person influences on metacomprehension accuracy, but also showed that between-person variability may be a more important factor than within-person variability at study, but almost equally important at test. The largest between-person effects were for main ideas and multiple choice judgments at both study and test, whereas large within-person effects were seen for most measures but mostly at test. MLM was less encouraging than tau-c about relative accuracy for predictions but more encouraging for postdictions. The power of MLM is that it separates between- and within-person influences on accuracy and can test cross-level interactions of these, so conclusions drawn from it may be the most accurate. MLM had the added of advantage of being able to reveal that both individual differences in average comprehension scores and average judgments are significant sources of variance in comprehension scores.

It is important to consider how metacomprehension as measured by summaries compared to metacomprehension as measured by multiple choice scores. Effect sizes for absolute and relative accuracy were similar for the two dependent measures at both study

and test and these were comparably affected by age. MLM suggested less comparable between-person and within-person accuracy. MLM showed that at study, relative accuracy on overall summary quality is stronger than it is for multiple-choice and between-person accuracy is stronger for multiple-choice, but this becomes roughly equalized at postdiction. MLM also suggests slightly more age similarity for judgments of overall summary quality than of multiple choice scores. Thus, although measuring metacomprehension using summaries may not result in wildly different conclusions than measuring with multiple choice scores, MLM suggests that there are some differences that should probably be taken into consideration. If one is concerned with particular aspects of summaries (e.g. themes, important details), rather than its overall quality, then one might want to consider the differences among measurement approaches more seriously.

There are some limitations inherent in the study of metacomprehension that are not present in the study of metamemory. First, people vary in their understanding of what it means to summarize (Brown & Day, 1983; Kintsch & van Dijk, 1978) and these beliefs may override explicit instructions given during the experiment about how to summarize. Second, there may be uncertainty about what constitutes a main idea or how main ideas would be identified by the experimenters, and this uncertainty may have created noise in participants' ability to make accurate judgments. Third, participants may have had a harder time discriminating among themes and percentage of main ideas across passages because of a restriction of range. Fourth, the size of some effects may be over-estimated because of a disconnect between what participants and researchers believe to be worthy of summary inclusion and because participants were told they would be penalized for the

inclusion of unimportant details. More details may be outputted when the instructions are to free recall rather than to summarize (van Dijk, 1979) and participants may have remembered more details than they included in their summaries because of beliefs about what should be included or how they would be scored. Fifth, within-person metacomprehension accuracy may be improved by requiring participants to make pre-judgments attempt at recall (Dunlosky, Rawson, & Middleton, 2005). All of these should be taken into account in future studies of metacomprehension using text summaries.

Is Metacomprehension Affected by Age, Reading Goal, Rereading, and Working Memory?

The effect of age on metacomprehension was inconsistent across dependent measures and as a function of statistical analysis. Bias scores showed more overconfidence for older adults than younger adults on overall summary quality and multiple choice. The only main effect of age on relative accuracy as measured by tau-c was that younger adults were better than older adults at judging the relative number of main ideas among their summaries at test. This is mostly consistent with the metamemory literature which finds that older adults tend to be more overconfident than younger adults but that age equivalency is usually found in relative accuracy (Hertzog & Hultsch, 2000). MLM, on the other hand, indicated age equivalency in both person- and passage-level judgments for main idea and overall summary metacomprehension (predictions and postdictions) but a robust younger adult advantage judging important details. MLM also showed that age had some influence on multiple choice judgments, but not independent of rereading effects. So, whereas MLM suggests that both between and within-person metacomprehension ability are likely fairly comparable between younger and older adults, other than older adults struggling to monitor details, tau-c and bias suggest more

age equivalency for within-person metacomprehension ability (tau-c) than for between person ability (bias).

Some of the inconsistency in age effects across various measures of metacomprehension may be due to age differences in beliefs about what constitutes "quality" in an overall summary or some of the components of a summary. However, older adults' overconfidence about their multiple choice scores suggests that this cannot fully account for the results. Unlike metamemory where some age differences have been attributed to differences in episodic memory and not monitoring (Connor, Dunlosky, & Hertzog, 1997), age differences in metacomprehension do not appear to be simply a function of age differences in comprehension; there were also age differences in judgment magnitude on all dependent measures except themes and overall summary quality. Although older adults' judgments and comprehension were almost always lower than those of younger adults, sometimes their metacomprehension was comparable to younger adults' and other times it was worse.

The effect of reading goal also varied as a function of statistical analysis. According to bias and tau-c, reading goal had no main effect with the exception of better theme discrimination by those in the acquaintance condition. MLM, however, indicated that the professor/boss condition showed superiority, mostly at the person-level, for all measures except multiple choice scores. Furthermore, MLM did not reveal any interactive effects of age and reading goal, but bias and tau-c did, although in opposite ways. Bias showed that younger adults tended to be inferior for acquaintances relative to professors and older adults tended to be inferior for bosses relative to acquaintances, but tau-c indicated that younger adults tended to be inferior for professors and older adults

for acquaintances. The effect of reading goal on overconfidence suggests that younger adults may be more at risk for spreading misinformation socially, but older adults may be more at risk for misrepresenting their comprehension in professional situations. It also suggests that there may be age differences in the standards held for different audiences—younger adults may set a higher standard for professors than acquaintances and older adults a higher standard for acquaintances than bosses. The pattern of relative accuracy points to a more serious issue for younger adults who are often students who need to be able to judge the relative quality of their comprehension so that under time constraints they can make strategic decisions about where to allocate further study. Then again, older adults who are retired spend more time summarizing for acquaintances than bosses, so their behavior poses a potentially significant risk for social misinformation exchange.

Different statistical analyses also supported divergent inferences about the effects of rereading and working memory. The influence of rereading and working memory on bias and tau-c were relatively minor. MLM analysis, in contrast, showed that rereading and working memory capacity influence all measures of metacomprehension. Rereading affected both within- and between-person accuracy, with some benefit for theme and important detail judgments but mixed effects on judgments of main ideas, overall summary quality, and multiple choice scores. Rereading tended to be a sign of struggle, generally, but when employed by those with higher working memory it appeared to provide a benefit. Broadly speaking, higher working memory tended to provide an advantage and mediated almost half of the effects of rereading on metacomprehension.

One might expect that simultaneously reading for comprehension and monitoring one's comprehension would be taxing on working memory. Consistent with that

expectation, working memory capacity significantly predicted both comprehension and, according to MLM primarily, metacomprehension. We might also expect that because older adults have worse working memory, they might have a harder time monitoring their comprehension. Their comprehension was worse and by some measures their metacomprehension was as well. But bias and tau-c did not implicate rereading as a significant factor mediating age differences (when they were found). MLM, on the other hand, suggested that younger adults with higher working memory capacity use rereading to excel but that older adults who reread achieved similar or worse outcomes than younger adults. This seems consistent with prior research that found that older adults with lower working memory capacity were worse at predicting their text learning (Miles & Stine-Morrow, 2004) and monitoring text inconsistencies that were farther apart in texts (Zabrucky et al., 1993). But it also indicates that age and working memory effects on metacomprehension depend on the specific dependent measure.

Some of the rereading and working memory effects in this study are inconsistent with previous literature. Prior research found that rereading of passages in full has been shown to improve metacomprehension (Dunlosky & Rawson, 2005; Rawson et al., 2000) but sometimes only for those with low working memory capacity (Griffin et al., 2008). This study shows that that may not be the case when self-directed, selective rereading is afforded. Rereading may have allowed those with lower working memory capacity to perform better than they otherwise would have, but rereading seemed to provide the largest benefit for those with higher working memory capacity. Older adults were previously shown to be less likely than younger adults to reread when they detect inconsistencies (Zabrucky & Moore, 1994) or when asked to prioritize accuracy over

speed (Stine-Morrow et al., 2006), but the reason is unknown. I found evidence that older adults reread more than younger adults and that there was no demosntrable benefit to this. Older adults may be laboring in vain or they may only be able to use rereading to achieve equivalency with younger adults in some cases. The pattern of effects of working memory and rereading on metacomprehension, combined with the age differences in text ratings, suggests that metacomprehension measured with text summaries is a relatively resource-demanding task for which domain familiarity may not be able to compensate.

Self-Reported Metacomprehension

In this study, I observed a change in self-reported metacomprehension and reading self-efficacy that has not previously been reported. After extended experience with summarizing and judging summaries, participants reported lower metacomprehension and lower reading self-efficacy than they had at the start of the experiment. Although reports on both measures were significantly correlated with comprehension scores, there was a stronger relationship of self-reported metacomprehension than reading self-efficacy to comprehension scores. That people can be induced to perceive their reading comprehension abilities more accurately after attempting to judge their summaries of expository texts is encouraging for metacomprehension training.

There were no age differences in reading self-efficacy, but younger adults reported lower metacomprehension than older adults. This is in contrast to earlier studies showing age equivalency in metacomprehension (Lin et al., 2000) and lower reported metacomprehension by older adults (De Beni et al., 2007; Moore et al., 1997a) than younger adults. It is unclear why younger adults would report lower metacomprehension

than older adults, particularly because their self-perception is inconsistent with their higher comprehension scores and their demonstrated age equivalency or superiority in metacomprehension as shown by other measures. Younger adults may set a higher comprehension standard for themselves, believe they are competent but not superior for their age group, or estimate a higher standard from the experimenter than older adults. These explanations are particularly viable given that the younger adults were sampled from a highly competitive and demanding university.

Despite showing reduced comprehension relative to younger adults, older adults reported more confidence in their comprehension capacity, more use of regulatory strategies, and an equivalent understanding of comprehension processes on the MCS. Older adults also reported equivalent reading self-efficacy and judged the passages to be less difficult than younger adults. Fortunately, older adults also reported greater valuing of comprehension than younger adults, so it may be possible to train older adults to more accurately assess their text comprehension. Older adults, however, showed a decrease in the extent to which they believe comprehension is under their control, so there would likely be some resistance to training.

Conclusions

Prior experiments on age differences in metacomprehension mostly concluded that there is age equivalency in metacomprehension (Dunlosky et al., 2006; Moore & Zabrucky, 1992; Olin & Zelinski, 1997; Zabrucky & Moore, 1994; Zabrucky et al., 1987) although occasionally age-related deficits have been reported (Baker et al., 2010; Miles & Stine-Morrow, 2004; Zabrucky et al., 1993). Most of these studies looked at the ability to detect inconsistencies in text (Moore & Zabrucky, 1992; Zabrucky & Moore, 1994;

Zabrucky et al., 1987), only two studies looked at the ability to predict comprehension test performance (Dunlosky et al., 2006; Olin & Zelinski, 1997), and only one of these studies (Baker et al., 2010) looked at age differences in postdiction accuracy. Given these mixed findings, and in comparing them to the present results which vary depending on the statistical measure used, in short, measurement matters—the paradigm, the dependent variables, and the statistical analyses all matter. According to traditional statistical measures, older adults' metacomprehension weaknesses compared to younger adults were with overall summary quality, multiple-choice, and to some extent main ideas. According to MLM, however, older adults showed more equivalency with younger adults on these measures but robust weakness with important detail judgment. Overall, results from the present study are fairly consistent with prior research, pointing to a mix of age equivalency and older adult deficits in metacomprehension, but this study builds on previous research in a several ways. It showed that according to statistical measures traditionally used to measure metacomprehension older adults tend to be more overconfident on average than younger adults but are similar to younger adults in their ability to discriminate among passages more or less well understood. In contrast, according to MLM, the effect of age on metacomprehension is as much, if not more, a function of the comprehension measure (e.g. main ideas vs. important details) as the source of the variability (within-person vs. between person).

Given how metacomprehension has traditionally been measured, it is important to consider more closely how participants' metacomprehension examined with summaries as the criterion compares to their metacomprehension examined with multiple-choice test performance. Effect sizes for age differences in bias were relatively small for overall

summary bias and multiple choice bias, pre and post, but they were larger overall for multiple choice bias. Notably, the largest age difference was for multiple-choice postdiction in which only older adults were still significantly overconfident. In all other cases, both younger and older adults were significantly overconfident. It is not clear whether it is better to measure comprehension and metacomprehension with summaries or multiple choice tests, and the truth is that they measure slightly different constructs. But, to the extent that they tap the same construct, summary measurement paints a slightly more optimistic picture for aging.

Given the small effect sizes for metacomprehension judgment accuracy and the general pattern of overconfidence displayed by participants, we should probably be encouraged by the reduction in MCS and RSE scores from pre to post administrations of the surveys. People can learn to reduce overconfidence simply by being asked to reflect on their comprehension. Nevertheless, it is a concern that older adults report more confidence than younger adults when older adults' comprehension was actually worse.

Further studies should attempt to account for age differences in comprehension and metacomprehension. Working memory capacity may account for some of the age-related differences and this can be explored by centering predictors at the grand mean rather than within age group. However, even with grand mean centering of predictors, tests of cross-sectional age mediation can be problematic (Lindenberger, von Oertzen, Ghisletta, & Hertzog, 2011). Consequently, working memory load should be experimentally manipulated within a cross-sectional design and longitudinal studies of working memory change and comprehension change are needed (Hultsch, Hertzog, Dixon, & Small, 1998). Another possible explanation is that persons of different age

groups have different concepts of what constitutes high quality for different aspects of a summary or for the summary overall. Another possibility is that persons of different age groups set different standards for comprehension and perhaps summarizing specifically, even for different audiences. Future studies could also examine motivational influences on decisions about whether to reread, how strategic decisions to reread for specific tasks may be, and how working memory interacts with other variables to influence whether rereading will prove worthwhile.

Overall, this dissertation found that as measured with text summaries, older adult metacomprehension is similar to that of younger adults, despite older adults showing worse text comprehension beyond the gist level. Both younger and older adults stand to improve their comprehension and metacomprehension abilities, as assessed with text summaries. Fortunately, a task such as the one used in this study appears be one way to induce more accurate self-perceptions of text comprehension abilities and could serve as the methodological basis for future metacomprehension studies and training development.

APPENDIX A: SAMPLE PASSAGE AND QUESTIONS

Television Newscasts

Relaying information and images instantly, television newscasts have allowed viewers to form their own opinions about various political events and political leaders. In many instances, television newscasts have even fostered active dissent from established government policies. It is no coincidence that, in the 1960's, the civil rights movement took hold in the United States with the advent of television, which was able to convey both factual information and such visceral elements as outrage and determination. Only when all of America could see, on the nightly newscasts, the civil disobedience occurring in places like Selma and Montgomery did the issue of civil rights become a national concern rather than a series of isolated local events. By relaying reports from cities involved to an entire nation of watchers, television showed viewers the scope of the discontent and informed the disenfranchised that they were not alone. The ability of television news to foster dissent has also been affected by increasingly widespread access to video cameras, so that the news presented on television now comes from the bottom up as well as from the top down. Across the world, dissidents have used video equipment to gather visual evidence of human rights abuses. Uncensored images and information have then been transmitted across otherwise closed borders by television newscasts. One professor of popular culture, Jack Nachbar, views the personal video camera as a "truth-telling device that can cut through lies." That claim presumes, though, that the television viewer can believe what he or she sees. But the motivation of the photographer must be taken into account, and the videotape that appears on television can, like still photography, be staged and even faked. When and if propagandists for some government utilize computer-generated effects, viewers will have more trouble believing what they see. However, even if seeing is not automatically believing, at least seeing is seeing--and in some repressive regimes, seeing is the fastest road to freedom.

1. The passage is primarily concerned with ways in which

a) television newscasts deliberately distort information
b) television affects viewers by its presentation of news
c) truth frustrates efforts by the media to constrain it
d) viewers of television newscasts cannot sort out fact from fiction
e) governments manage to control television newscasts

2. Which of the following, if true, would most strengthen the assertion about television and the American civil rights movement?

a) Many filmed reports of civil disobedience were censored by television executives during the 1960s
b) Recent studies have questioned the objectivity with which television newscasts presented reports of civil disobedience during the 1960s
c) A biography of a major civil rights leader describes in detail the occasions on which the leader was featured in television newscasts in the 1960s
d) A 1960s poll shows that those Americans who considered civil rights a national priority had seen television newscasts of civil disobedience
e) Many of the reporting techniques used today originated in newscasts covering the 1960s civil rights movement

3. It can be inferred from the passage that television newscasts would be better at informing public opinion if

a) newscasts presented only competing views and not one-sided views
b) personal videos were banned from television newscasts
c) technology was developed to detect when videos had been tampered with
d) highly visceral information were not presented during television newscasts
e) only factual information were presented during television newscasts

4. The author suggests a major reason why television newscasts are effective at influencing public opinion. Based on this argument, which medium below would be the most effective at influencing public opinion?

a) daily newspapers
b) radio broadcasts
c) classroom instruction
d) grassroots movements based on word of mouth
e) witnessing newsworthy events first hand

5. According to the passage, television coverage of the civil rights movement did all of the following EXCEPT

a) inform dissenters that they were not alone

b) convey factual information
c) present emotional elements such as anger
d) portray the scope of the dissent
e) express opinions of the political leaders

6. Jack Nachbar, who is quoted in the passage, is

a) a popular culture professor
b) a government propagandist
c) a reporter for a professional news agency
d) a civil rights activist
e) a prominent political figure

7. The author explicitly states that the believability of television news may be compromised by

a) effects produced by computers
b) videos from personal cameras
c) photographers for professional news agencies
d) established government policies
e) reports that are transmitted across closed borders

8. The passage states that when nightly newscasts portrayed civil dissent in the 1960s,

a) it incited dissent in places like Selma and Montgomery
b) it created a national concern for civil rights
c) it started a series of isolated local events
d) viewers formed opinions about political leaders
e) interest in personal video cameras increased

APPENDIX B: SCORING RUBRIC

Television Newscasts

____ Theme: TV has affected viewers' politics, but the accuracy of the information they provide is questionable.

_____Main Idea 1: TV changed the way information was experienced

____ ID 1: transmits info quickly/efficiently

____ Main Idea 2: TV has affected(impacted) politics/culture/social issues

____ ID 1: role in the civil rights movement

____ ID 2: can be emotional/provocative/uncensored

____ ID 3: transmits info widely/bridges people

____ ID 4: personal cameras/bottom up transmission

____ Main Idea 3: TV newscasts can be untruthful/unreliable

____ ID 1: may be biased/motivated/used for propaganda

____ ID 2: can be staged or faked

Overall Summary Score ____

Real vs. Fake Art

____ Theme: Fake art is generally seen as inferior to real art (criticized), but it deserves more credit (has value/as much as the original).

____ Main Idea 1: Fake art is not liked as much as real art

____ ID 1: not a relic/connected to past/history/historical context

____ ID 2: not work of genius/original artist/authenticity

____ ID 3: cheats purchasers of money/status

____ ID 4: doesn't elicit reverence/spiritual refreshment/positive feelings/sentimentality

____ ID 5: make scientific dating harder

____ ID 6: deform understanding of the past

____ Main Idea 2: Fake art has value

____ ID 1: show that talented artists still exist/fake artist as good as original artist

____ ID 2: help develop methods of analysis

____ ID 3: raise questions about aesthetics (appearance and reactions to appearance)

____ Main Idea 3: There are pros and cons of fake art.

Overall Summary Score _____

Precision of Science

____ Theme: Science is not always precise and often involves luck.

____ Main Idea 1: Scientific theories are not always precise/accurate/correct about details

____ ID 1: Dalton and atomic theory/atomic weights

____ ID 2: Copernicus and heliocentric universe/planetary orbits/not Earth-centered

____ Main Idea 2: Luck plays a role in science

____ ID 1: Pasteur and yeast/not hay bacillus/boiling/not heat resistant/origin of life

____ ID 2: Mendel and genetics/traits/pea plants/hybrids/mathematical relationships/plant interbreeding

Overall Summary Score _____

Women in the Workplace

____ Theme: Although women's roles in the 19th century were primarily seen as domestic, many women entered the paid workplace.

____ Main Idea 1: More women entered the paid workplace during the 19th century

____ Main Idea 2: Women entered the paid workplace (or just "worked") out of economic necessity

____ ID 1: young/unmarried

____ ID 2: husband unable to work

____ ID 3: widowed

____ Main Idea 3: Women's role was seen as domestic at the time

____ ID 1: earning ability was secondary/should be supported, not support

____ Main Idea 4: Women in the paid workplace were not treated fairly/seen as equals

____ ID 1: paid less than men

____ ID 2: couldn't advance/seen as casual laborers

Overall Summary Score _____

Zoo Habitats

____ Theme: Despite the good intentions of zoos, animals do not generally thrive in them.

____ Main Idea 1: Zoos try to save animals from extinction.

____ ID 1: save animals from destroyed habitat/deforestation

____ ID 2: save animals from inbreeding.

____ Main Idea 2: Zoos try to better the lives of animals in zoos

____ ID 1: use interior decorating/optical illusions

____ ID2: attend to psychological health (mindset; sensibilities)/stress of animals

____ ID3: attempts to better understand life in wild/differences between zoo and wild life/what animals need

____ Main Idea 3: Zoo animals are different from wild animals

____ ID1: no energy release/stuck in cages/pacing

____ ID 2: lose/atrophy wild skills

____ ID 3: no natural culture/social structure/live in vacuum

____ ID 4: lazy/bored

_____ID 5: dependent on humans/don't search for own food

____ Main Idea 4: Zoos may be changing species over generations

Overall Summary Score _____

American Indians

____ Theme: American Indians did not have the political/societal structure (hierarchy) that settlers did, and the settlers reacted to this in varying ways.

____ Main Idea 1: American Indian society was not structured hierarchically (no strong leadership or coercive politics)

____ ID 1: usually a council ruled

____ ID 2: chieftain was an honorary (symbolic) position/position of respect

____ ID 3: chiefs had ceremonial/religious roles (not political/economic ones)

____ ID 4: Indian society was based on equality/kinship (no social classes/no private property)

____ Main Idea 2: Europeans tried to change the American Indian system

____ ID 1: Europeans tried to force the role of chief/impute monarchy/make less democratic

____ ID 2: need for trade/treaties had a role in the change

____ Main Idea 3: Positive of the American Indian society were noticed/recognized/accepted by some settlers

Overall Summary Score _____

APPENDIX C: SCORING RULES

General Rules of Thumb

1. Always anchor with the rubric first. Pick an item from the rubric and then look for it in the transcription. Indicate on the checklist whether the essence of that item was in the transcription.
2. Themes, main ideas, and important details only need to be *gist consistent* to earn credit. Ask yourself, "Does the participant capture the *essence* of the concept in our rubric?"
    - Although you are looking for gist consistency, be careful about assuming too much.
        - "Almost" is not the same as "gist". If someone *almost* captures the essence, that is not the same as actually capturing the essence.
    - Don't be distracted by a participant's use of a keyword. If they don't use it in a way that shows understanding of a theme, main idea, or important detail, they don't get credit for it.
    - If you are unsure whether to give credit, ask yourself 1) could the participant have meant something else? 2) if so, is it highly unlikely they meant this other thing? 3) if so, then there is not reasonable doubt so give them credit.
    - If you are rereading too many times and are still unsure whether to give credit, don't give credit.
3. You can give a participant credit for an important detail if it is more specific than the detail in our rubric, but not if it is more general.
4. You can give a participant credit for a main idea regardless of its specificity, as long as it is gist consistent and as long as you don't base the credit on the same information on which you base credit for an important detail.
5. If you can parse a sentence or phrase so that the participant can get credit for more than one item, do so.
6. Do not double score. If you use a phrase in a transcription to provide evidence for an important detail that phrase alone cannot also be the sole evidence for a main idea.
7. Elaboration—information that was not in the passage but does not contradict the passage—does not get scored.
8. Double check your scores. Make sure you can provide evidence for the credit you give.

Themes Scoring

In general, if a theme meets threshold for a score but also contains an inaccuracy, bump it down one score.

X = missing theme (only transcriber note indicating no theme)

1= no theme (participant is explicit about not being able to generate a theme)

2= incorrect theme (any incorrect theme even if it doesn't contradict the passage)

3= subject matter (topic) only

4= overly general theme (accurately identifies the debate/central question of the passage or makes an accurate point about the topic but does not meet threshold for a higher score)

5= half the theme is present in at least gist form, but there is nothing additional (or the additional information is incorrect)

6=half of the theme is present in at least gist form and there is additional information which is accurate but does not meet threshold for gist consistency with other half of theme

7= both parts of the theme are present in at least gist form

Overall Summary Scoring

In order of important, consider the presence of the following. Also consider the distribution of MI and ID—it is better to have more even distribution than one main idea with all of its important details:

1. Main Ideas
2. Important Details
3. Organized/Coherent
4. Efficient
5. Inaccuracies

7 (5/5)	Demonstrates Superiority	• Contains all (or almost all) main ideas and important details • Organized and coherent • Efficient • No inaccuracies, opinions, previous knowledge, or unimportant details
6 (4/5)	Suggests Superiority	• Contains most main ideas and several important details • Organized and coherent, • Relatively efficient • No inaccuracies, opinions, previous knowledge, or unimportant details
5 (3/5)	Demonstrates Competence	• Contains some main ideas and some details • Shows some organization and coherence • Relatively efficient • Minimal inaccuracies, opinions, previous knowledge, or unimportant details
4 (2-3/5)	Suggests Competence	• Contains mostly main ideas or mostly important details • Show some organization and coherence • Relatively efficient • Minimal inaccuracies, opinions, previous knowledge, or unimportant details
3 (2/5)	Suggests Incompetence	• Contains only one main idea and only 1-2 important details • Organization and coherence are limited at best • Relatively inefficient or too lacking in information to be inefficient • May contain one or more inaccuracies
2 (1/5)	Demonstrates Incompetence	• Contains no main ideas and very few important details

		• Lacking in organization, coherence, or efficiency or is too short to judge • May contain one or more inaccuracies
1 (0/5)	Incompetent	• Contains no main ideas or important details • If anything is present it is only opinion or nonsensical • May contain one or more inaccuracies

REFERENCES

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40*(1), 278-289. doi: 10.3758/brm.40.1.278

Adams, C. (1991). Qualitative age differences in memory for text: A life-span developmental perspective. *Psychology and Aging, 6*(3), 323-336. doi: 10.1037/0882-7974.6.3.323

Algina, J. & Swaminathan, H. (2011). Centering in two-level nested designs. In J.J. Hox & J.K. Roberts (Eds.), Handbook for advanced multilevel analysis. (pp. 285-312). New York, NY US: Routledge/Taylor & Francis Group.

Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica, 128*(1), 110-118. doi: 10.1016/j.actpsy.2007.10.006

Baker, J., Dunlosky, J., & Hertzog, C. (2010). How Accurately Can Older Adults Evaluate the Quality of Their Text Recall? The Effect of Providing Standards on Judgment Accuracy. *Appl Cogn Psychol, 24*(1), 134. doi: 10.1002/acp.1553

Board, T. C. (1997). 10 real SATs. New York: College Entrance Examination Board.

Borella, E., Ghisletta, P., & Ribaupierre, A. d. (2011). Age differences in text processing: The role of working memory, inhibition, and processing speed. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 66B*(3), 311-320. doi: 10.1093/geronb/gbr002

Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension *Visual information processing.* Oxford England: Academic.

Brébion, G. (2003). Working Memory, Language Comprehension, and Aging: Four Experiments to Understand the Deficit. *Experimental Aging Research, 29*(3), 269-301. doi: 10.1080/03610730303725

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning & Verbal Behavior, 22*(1), 1-14. doi: 10.1016/s0022-5371(83)80002-4

Bryan, J., & Luszcz, M. A. (1999). Speed of information processing and working memory as mediators of age differences in prose recall. *Australian Psychologist, 34*(2), 132-137. doi: 10.1080/00050069908257442

Byrd, M. (1985). Age differences in the ability to recall and summarize textual information. *Experimental Aging Research, 11*(2), 87-91. doi: 10.1080/03610738508259285

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D.Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 31-68). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Chiang, E. S., Therriault, D. J., & Franks, B. A. (2010). Individual differences in relative metacomprehension accuracy: Variation within and across task manipulations. *Metacognition and Learning, 5*(2), 121-135. doi: 10.1007/s11409-009-9052-6

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12*(1), 50-71. doi: 10.1037/0882-7974.12.1.50

De Beni, R., Borella, E., & Carretti, B. (2007). Reading comprehension in aging: the role of working memory and metacomprehension. [Comparative Study]. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 14*(2), 189-212. doi: 10.1080/13825580500229213

DeDe, G., Caplan, D., Kemtes, K., & Waters, G. (2004). The Relationship Between Age, Verbal Working Memory, and Language Comprehension. *Psychology and Aging, 19*(4), 601-616. doi: 10.1037/0882-7974.19.4.601

Dijkstra, K. (2001). Old readers: Slow readers or expert readers? In D. Schram & G. Steen (Eds.), *The psychology and sociology of literature: In honor of Erlund Ibsch.* (pp. 87-105). Amsterdam Netherlands: John Benjamins Publishing Company.

Dudukovic, N. M., Marsh, E. J., & Tversky, B. (2004). Telling a Story or Telling it Straight: The Effects of Entertaining Versus Accurate Retellings on Memory. *Appl Cogn Psychol, 18*(2), 125-143. doi: 10.1002/acp.953

Dunlosky, J., Baker, J. M. C., Rawson, K. A., & Hertzog, C. (2006). Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. *Psychology and Aging, 21*(2), 390-400. doi: 10.1037/0882-7974.21.2.390

Dunlosky, J., & Rawson, K. A. (2005). Why Does Rereading Improve Metacomprehension Accuracy? Evaluating the Levels-of-Disruption Hypothesis for the Rereading Effect. *Discourse Processes, 40*(1), 37-55. doi: 10.1207/s15326950dp4001_2

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language, 52*(4), 551-565. doi: 10.1016/j.jml.2005.01.011

Durso, F. T., Rawson, K. A., & Girotto, S. (2007). Comprehension and situation awareness. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandowsky & T. J. Perfect (Eds.), *Handbook of applied cognition (2nd ed.).* (pp. 164-193). Hoboken, NJ US: John Wiley & Sons Inc.

Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., & Dooley, S. (2005). Summary street®: Computer support for comprehension and writing. *Journal of Educational Computing Research, 33*(1), 53-80. doi: 10.2190/dh8f-qjwm-j457-fqvb

Freebody, P., & Anderson, R. C. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly, 18*(3), 277-294. doi: 10.2307/747389

Garner, R. (1985). Text summarization deficiencies among older students: Awareness or production ability? *American Educational Research Journal, 22*(4), 549-560. doi: 10.2307/1163139

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15*(1), 84-93.

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36*(1), 93-103. doi: 10.3758/mc.36.1.93

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*(4), 339-387. doi: 10.1006/cogp.2001.0769

Harker, J. O., Hartley, J. T., & Walsh, D. A. (1982). Understanding discourse: A life-span approach. *Advances in Reading/Language Research, 1*, 155-202.

Hartley, J. T. (1986). Reader and text variables as determinants of discourse memory in adulthood. *Psychology and Aging, 1*(2), 150-158. doi: 10.1037/0882-7974.1.2.150

Hartley, J. T. (1993a). Aging and individual differences in memory for written discourse. In L. L. Light & D. M. Burke (Eds.), *Language, memory, and aging.* (pp. 36-57). New York, NY US: Cambridge University Press.

Hartley, J. T. (1993b). Aging and prose memory: Tests of the resource-deficit hypothesis. *Psychology and Aging, 8*(4), 538-551. doi: 10.1037/0882-7974.8.4.538

Hartley, J. T., Stojack, C. C., Mushaney, T. J., Annon, T. A. K., & Lee, D. W. (1994). Reading speed and prose memory in older and younger adults. *Psychology and Aging, 9*(2), 216-223. doi: 10.1037/0882-7974.9.2.216

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition (2nd ed.).* (pp. 417-466). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Hess, T. M. (1995). Aging and the impact of causal connections on text comprehension and memory. *Aging & Cognition, 2*(3), 216-230. doi: 10.1080/13825589508256599

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging, 24*(2), 462-475. doi: 10.1037/a0014417

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39*(1), 101-117. doi: 10.3758/bf03192848

http://audacity.sourceforge.net/.

Hultsch, D. F., Hertzog, C., & Dixon, R. A. (1990). Ability correlates of memory performance in adulthood and aging. *Psychology and Aging, 5*(3), 356-368. doi: 10.1037/0882-7974.5.3.356

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory change in the aged*. New York, NY US: Cambridge University Press.

Ikeda, K., & Kitagami, S. (2012). The effect of working memory capacity and mental effort on monitoring accuracy in text comprehension. *Psychologia: An International Journal of Psychological Sciences, 55*(3), 184-193. doi: 10.2117/psysoc.2012.184

Ikeda, K., & Kitagami, S. (2013). The interactive effect of working memory and text difficulty on metacomprehension accuracy. *Journal of Cognitive Psychology, 25*(1), 94-106.

Jackson, J. D., & Kemper, S. (1993). Age differences in summarizing descriptive and procedural texts. *Experimental Aging Research, 19*(1), 39-51. doi: 10.1080/03610739308253922

Jeong, H., & Kim, H. S. (2009). Aging and text comprehension: Interpretation and domain knowledge advantage. *Educational Gerontology, 35*(10), 906-928. doi: 10.1080/03601270902834601

Johnson, R. E. (2003). Aging and the remembering of text. *Developmental Review, 23*(3), 261-346. doi: 10.1016/s0273-2297(03)00009-1

Kellogg, R. T. (2007). Are written and spoken recall of text equivalent? *The American Journal of Psychology, 120*(3), 415-428.

Kendeou, P., Bohn-Gettler, C., & Fulton, S. (2011). What we have been missing: The role of goals in reading comprehension. In M. T. McCrudden, J. P. Magliano & G. Schraw (Eds.), *Text relevance and learning from text.* (pp. 375-394). Charlotte, NC US: IAP Information Age Publishing.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182. doi: 10.1037/0033-295x.95.2.163

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294-303. doi: 10.1037/0003-066x.49.4.294

Kintsch, W., & Kozminsky, E. (1977). Summarizing stories after reading and listening. *Journal of Educational Psychology, 69*(5), 491-499. doi: 10.1037/0022-0663.69.5.491

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394. doi: 10.1037/0033-295x.85.5.363

Lin, L.-M., Moore, D., & Zabrucky, K. M. (2000). Metacomprehension knowledge and comprehension of expository and narrative texts among younger and older adults. *Educational Gerontology, 26*, 737-749.

Lindenberger, U., von Oertzen, T., Ghisletta, P., & Hertzog, C. (2011). Cross-sectional age variance extraction: What's change got to do with it? *Psychology and Aging 26*(1), 34-47. doi: 10.1037/a0020525.

Linderholm, T., Cong, X., & Zhao, Q. (2008). Differences in low and high working-memory capacity readers' cognitive and metacognitive processing patterns as a function of reading for different purposes. *Reading Psychology, 29*(1), 61-85. doi: 10.1080/02702710701568587

Linderholm, T., Kwon, H., & Wang, X. (2011). The effects of reading purpose on advanced readers. In M. T. McCrudden, J. P. Magliano & G. Schraw (Eds.), *Text relevance and learning from text.* (pp. 197-222). Charlotte, NC US: IAP Information Age Publishing.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94*(4), 778-784. doi: 10.1037/0022-0663.94.4.778

Linderholm, T., & Wilde, A. (2010). Students' beliefs about comprehension when reading for different purposes. *Journal of College Reading and Learning, 40*(2), 7-19.

Linderholm, T., & Zhao, Q. (2008). The impact of strategy instruction and timing of estimates on low and high working-memory capacity readers' absolute monitoring accuracy. *Learning and Individual Differences, 18*(2), 135-143. doi: 10.1016/j.lindif.2005.09.002

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2000). *SAS system for mixed models* (4th ed.). Cary, NC: SAS Institute, Inc.

Liu, C.-j., Kemper, S., & Bovaird, J. A. (2009). Comprehension of health-related written materials by older adults. *Educational Gerontology, 35*(7), 653-668. doi: 10.1080/03601270902885504

Lorch, R. F., Lorch, E. P., & Klusewitz, M. A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology, 85*(2), 239-252. doi: 10.1037/0022-0663.85.2.239

Madden, C. J., & Dijkstra, K. (2010). Contextual constraints in situation model construction: An investigation of age and reading span. *Aging, Neuropsychology, and Cognition, 17*(1), 19-34. doi: 10.1080/13825580902927604

Maki, R. H. (1998a). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition, 26*(5), 959-964. doi: 10.3758/bf03201176

Maki, R. H. (1998b). Test predictions over text material. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice.* (pp. 117-144). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 663-679. doi: 10.1037/0278-7393.10.4.663

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 609-616. doi: 10.1037/0278-7393.16.4.609

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels

of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43. doi: 10.1207/s1532690xci1401_1

Meyer, B. J. F., & Rice, G. E. (1989). Prose processing in adulthood: The text, the reader, and the task. In L. W. Poon, D. C. Rubin & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life.* (pp. 157-194). New York, NY US: Cambridge University Press.

Miles, J. R., & Stine-Morrow, E. A. L. (2004). Adult age differences in self-regulated learning from reading sentences. [Research Support, U.S. Gov't, P.H.S.]. *Psychol Aging, 19*(4), 626-636. doi: 10.1037/0882-7974.19.4.626

Miller, L. M. S., Stine-Morrow, E. A. L., Kirkorian, H. L., & Conroy, M. L. (2004). Adult Age Differences in Knowledge-Driven Reading. *Journal of Educational Psychology, 96*(4), 811-821. doi: 10.1037/0022-0663.96.4.811

Moore, D., & Zabrucky, K. (1992). Self-judged comprehension in adults: Effects of age and skill. *Experimental Aging Research, 18*(1-2), 3-7. doi: 10.1080/03610739208253904

Moore, D., Zabrucky, K., & Commander, N. E. (1997a). Metacomprehension and comprehension performance in younger and older adults. *Educational Gerontology, 23*(5), 467-475. doi: 10.1080/0360126970230506

Moore, D., Zabrucky, K., & Commander, N. E. (1997b). Validation of the Metacomprehension Scale. *Contemporary Educational Psychology, 22*(4), 457-471. doi: 10.1006/ceps.1997.0946

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology, 91*(3), 488-496. doi: 10.1037/0022-0663.91.3.488

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109-133. doi: 10.1037/0033-2909.95.1.109

Olin, J. T., & Zelinski, E. M. (1997). Age differences in calibration of comprehension. *Educational Gerontology, 23*(1), 67-77. doi: 10.1080/0360127970230106

Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., & Stine-Morrow, E. A. L. (2012). The effects of print exposure on sentence processing and memory in older adults: Evidence for efficiency and reserve. *Aging, Neuropsychology, and Cognition, 19*(1-2), 122-149. doi: 10.1080/13825585.2011.628376

Pierce, B. H., & Smith, S. M. (2001). The postdiction superiority effect in metacomprehension of text. *Memory & Cognition, 29*(1), 62-67.

Prat-Sala, M., & Redford, P. (2010). The interplay between motivation, self-efficacy, and approaches to studying. *British Journal of Educational Psychology, 80*(2), 283-305. doi: 10.1348/000709909x480563

Prat-Sala, M., & Redford, P. (2012). Writing essays: Does self-efficacy matter? The relationship between self-efficacy in reading and in writing and undergraduate students' performance in essay writing. *Educational Psychology, 32*(1), 9-20. doi: 10.1080/01443410.2011.621411

Radvansky, G. A. (1999). Aging, memory, and comprehension. *Current Directions in Psychological Science, 8*(2), 49-53. doi: 10.1111/1467-8721.00012

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American Journal of Psychology, 117*(2), 191-213. doi: 10.2307/4149022

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review, 14*(6), 1027-1042. doi: 10.3758/bf03193088

Radvansky, G. A., Gerard, L. D., Zacks, R. T., & Hasher, L. (1990). Younger and older adults' use of mental models as representations for text materials. *Psychology and Aging, 5*(2), 209-214. doi: 10.1037/0882-7974.5.2.209

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging, 16*(1), 145-160. doi: 10.1037/0882-7974.16.1.145

Ratner, H. H., Schell, D. A., Crimmins, A., Mittelman, D., & Baldinelli, L. (1987). Changes in adults' prose recall: Aging or cognitive demands? *Developmental Psychology, 23*(4), 521-525. doi: 10.1037/0012-1649.23.4.521

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 69-80. doi: 10.1037/0278-7393.28.1.69

Rawson, K. A., Dunlosky, J., & Theide, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition, 28*(6), 1004-1010.

Rayner, K., Juhasz, B. J., & Pollatsek, A. (2005). Eye Movements During Reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook.* (pp. 79-97). Malden: Blackwell Publishing.

Rayner, K., & Sereno, S. C. (1994). Eye movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics.* (pp. 57-81). San Diego, CA US: Academic Press.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164-171. doi: 10.1027/1015-5759/a000123

Riley, G. L., & Lee, J. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing, 13*(2), 173-189. doi: 10.1177/026553229601300203

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology, 8*(4), 535-543. doi: 10.1037/0894-4105.8.4.535

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime 2.0*. Pittsburgh: Psychology Software Tools.

Schommer, M., & Surber, J. R. (1986). Comprehension-monitoring failure in skilled adult readers. *Journal of Educational Psychology, 78*(5), 353-357. doi: 10.1037/0022-0663.78.5.353

Soederberg Miller, L. M. (2009). Age differences in the effects of domain knowledge on reading efficiency. *Psychology and Aging, 24*(1), 63-74. doi: 10.1037/a0014586

Souchay, C., & Isingrini, M. (2004). Age related differences in metacognitive control: Role of executive functioning. *Brain and Cognition, 56*(1), 89-99. doi: 10.1016/j.bandc.2004.06.002

Stine-Morrow, E. A. L., Loveless, M. K., & Soederberg, L. M. (1996). Resource allocation in on-line reading by younger and older adults. *Psychology and Aging, 11*(3), 475-486. doi: 10.1037/0882-7974.11.3.475

Stine-Morrow, E. A. L., Milinder, L.-A., Pullara, O., & Herman, B. (2001). Patterns of resource allocation are reliable among younger and older readers. *Psychology and Aging, 16*(1), 69-84. doi: 10.1037/0882-7974.16.1.69

Stine-Morrow, E. A. L., Noh, S. R., & Shake, M. C. (2010). Age differences in the effects of conceptual integration training on resource allocation in sentence processing. *The Quarterly Journal of Experimental Psychology, 63*(7), 1430-1455. doi: 10.1080/17470210903330983

Stine-Morrow, E. A. L., Shake, M. C., Miles, J. R., Lee, K., Gao, X., & McConkie, G. (2010). Pay now or pay later: Aging and the role of boundary salience in self-regulation of conceptual integration in sentence processing. *Psychology and Aging, 25*(1), 168-176. doi: 10.1037/a0018127

Stine-Morrow, E. A. L., Shake, M. C., Miles, J. R., & Noh, S. R. (2006). Adult age differences in the effects of goals on self-regulated sentence processing. *Psychology and Aging, 21*(4), 790-803. doi: 10.1037/0882-7974.21.4.790

Stine-Morrow, E. A. L., Soederberg Miller, L. M., Gagne, D. D., & Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging, 23*(1), 131-153. doi: 10.1037/0882-7974.23.1.131

Stine, E. A. L. (1990). On-line processing of written text by younger and older adults. *Psychology and Aging, 5*(1), 68-78. doi: 10.1037/0882-7974.5.1.68

Stine, E. A. L., & Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging, Neuropsychology, and Cognition, 1*(1), 2-16. doi: 10.1080/09289919408251446

Stuart, A. (1977). Spearman-like computation of Kendall's tau. *British Journal of Mathematical and Statistical Psychology, 30*(1), 104-112. doi: 10.1111/j.2044-8317.1977.tb00729.x

Taub, H. A., & Kline, G. E. (1976). Modality effects and memory in the aged. *Educational Gerontology, 1*(1), 53-60.

Taylor, B. M. (1984). The search for a meaningful approach to assessing comprehension of expository text. *National Reading Conference Yearbook, 33*, 257-263.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*(2), 129-160. doi: 10.1016/s0361-476x(02)00011-5

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505. doi: 10.3758/bf03192720

Van den Broek, P., Lorch, R. F., Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29*(8), 1081-1087. doi: 10.3758/bf03206376

Van der Linden, M., Hupet, M., Feyereisen, P., Schelstraete, M.-A., Bestgen, Y., Bruyer, R., . . . Seron, X. (1999). Cognitive mediators of age-related differences in language comprehension and verbal memory performance. *Aging, Neuropsychology, and Cognition, 6*(1), 32-55. doi: 10.1076/anec.6.1.32.791

van Dijk, T. A. (1979). Recalling and summarizing complex discourse. In W. Burghardt & K. Holke (Eds.), *Textverarbeitung/ Text Processing* (pp. 49-118). Berlin/New York: de Gruyter.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies in discourse comprehension*. New York: Academic Press.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journals of Gerontology, 48*(4), 157-171. doi: 10.1093/geronj/48.4.P157

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122*(3), 231-249. doi: 10.1037/0033-2909.122.3.231

Vieiro, P., & García-Madruga, J. A. (1997). An analysis of story comprehension through spoken and written summaries in school-age children. *Reading and Writing, 9*(1), 41-53. doi: 10.1023/a:1007932429184

Walczyk, J. J., & Taylor, R. W. (1996). How do the efficiencies of reading subcomponents relate to looking back in text? *Journal of Educational Psychology, 88*(3), 537-545. doi: 10.1037/0022-0663.88.3.537

Weaver, C. A. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(2), 214-222. doi: 10.1037/0278-7393.16.2.214

Weaver, C. A., & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition, 23*(1), 12-22. doi: 10.3758/bf03210553

Zabrucky, K., & Moore, D. (1994). Contributions of working memory and evaluation and regulation of understanding to adults' recall of texts. *Journals of Gerontology, 49*(5), 201-212. doi: 10.1093/geronj/49.5.P201

Zabrucky, K., Moore, D., & Schultz, N. R. (1987). Evaluation of comprehension in young and old adults. *Developmental Psychology, 23*(1), 39-43. doi: 10.1037/0012-1649.23.1.39

Zabrucky, K., Moore, D., & Schultz, N. R. (1993). Young and old adults' ability to use different standards to evaluate understanding. *Journals of Gerontology, 48*(5), P238-P244.

Zachary, R. A. (1986). Shipley Institute of Living Scale: Revised Manual. Los Angeles: Western Psychological Services.

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 920-933. doi: 10.1037/0278-7393.20.4.920